Computer Aided Detection of Microcalcification Clusters in Digital Mammogram Images

Kristin J. McLoughlin B.E.(Hons. I)

Department of Electrical and Computer Engineering

A thesis presented for the degree of
Doctor of Philosophy

University of Canterbury
Christchurch, New Zealand
February 10, 2004

Dance like no one is watching, love like you'll never be hurt, sing like no one is listening and live like it's heaven on earth.

William Purkey

If you don't go after what you want, you'll never have it. If you don't ask, the answer is always no. If you don't step forward, you're always in the same place.

Nora Roberts

Abstract

Recent advancements in computer technology have ensured that early detection of breast cancer, via computer aided detection (CAD) schemes, has become a rapidly expanding field of research. There is a desire to improve the detection accuracy of breast cancer without increasing the number of falsely identified cancers. The CAD scheme considered here is intended to assist radiologists in the detection of microcalcification clusters, providing a real contribution to the mammography screening process.

Factors that affect the detection accuracy of microcalcifications in digital mammograms include the presence of high spatial frequency noise, and locally linear high intensity structures known as curvilinear structures (CLS). The two issues considered are how to compensate for the high frequency image noise and how to detect CLS thus removing their influence on microcalcification detection. First, an adaptive approach to modelling the image noise is adopted. This is derived directly from each mammogram and is adaptable to varying imaging conditions. It is found that compensating for the high frequency image noise significantly improves microcalcification detection accuracy. Second, due to the varying size and orientation of CLS in mammogram images, a shape parameter is designed for their detection using a multiresolution wavelet filter bank. The shape parameter leads to an efficient way of distinguishing curvilinear structures from faint microcalcifications. This improves microcalcification detection performance by reducing the number of false positive detections related to CLS.

The detection and segmentation of microcalcification clusters is achieved by the development of a stochastic model, which classifies individual pixels within a mammogram into separate classes based on Bayesian decision theory. Both the high frequency noise model and CLS shape parameters are used as input to this segmentation process. The CAD scheme is specifically designed to be independent of the modality used, simultaneously exploiting the image data and prior knowledge available for microcalcification detection.

A new hybrid clustering scheme enables the distinction between individual and clustered microcalcifications, where clustered microcalcifications are considered more clinically sus-

picious. The scheme utilises the observed properties of genuine clusters (such as a uniform distribution) providing a practical approach to the clustering process. The results obtained are encouraging with a high percentage of genuine clusters detected at the expense of very few false positive detections.

An extensive performance evaluation of the CAD scheme helps determine the accuracy of the system and hence the potential contribution to the mammography screening process. Comparing the CAD scheme developed with previously developed microcalcification detection schemes shows that the performance of this method is highly competitive. The best results presented here give a sensitivity of 91% at an average false positive detection rate of 0.8 false positives per image.

Acknowledgements

First and foremost I would like to thank my supervisor Phil Bones for his time, effort, guidance and friendship. His enthusiasm and support has helped to make my PhD studies productive and enjoyable. I am also grateful to Richard Lane for his ideas and input during the writing of my thesis.

I would like to acknowledge Peter Kovesi from the Department of Computer Science & Software Engineering at the University of Western Australia for discussing his ideas and providing a real contribution to my work.

In the course of my research I enjoyed a short stay at the University Medical Centre Nijmegen, Nijmegen, The Netherlands, where I studied with a wonderful group of people in the radiology department. I am especially grateful to Nico Karssemeijer for allowing me to visit and ensuring my stay was both informative and enjoyable. Having the opportunity to study with one of the best in the mammography computer aided detection field was incomparable and considerably enhanced my research. I express my thanks to Jen Hendrick, who taught me about the mammography screening process and Ruben van Engen for showing me the direct digital mammography system and helping me to obtain direct digital mammogram data. In addition, my thanks goes to Celia Varela for showing me around Nijmegen and sharing many new and interesting experiences with me (including the Nijmegen night life).

I gratefully acknowledge the support of the Foundation for Research, Science & Technology. In particular the allowance for conference travel has been an invaluable part of the scholarship. I would also like to acknowledge the funding provided by the The Royal Society of New Zealand through the R.H.T. Bates postgraduate scholarship, and the William Georgetti scholarship awarded by the New Zealand Vice-Chancellors committee.

Finally, I would like to thank my family and friends for their support and love. You have all made my life extremely enjoyable and for that I am truly grateful.

Contents

Preface

Breast cancer is one of the most common forms of cancer among women. It would be safe to say that every person in New Zealand knows of someone who has been diagnosed with, or even worse died from this disease. The early detection and effective treatment of breast cancer is currently the only way of increasing the chance of survival for patients and consequently breast cancer screening programs have been introduced in a number of countries around the world. Due to the advancement in computer technology and digital image processing, the development of computer aided detection (CAD) schemes for the detection of breast cancer in digital mammography is becoming a rapidly expanding field of research. The aim of this thesis is to describe the development of a new CAD scheme for the detection of microcalcification clusters. The scheme is intended to assist radiologists in the detection of breast cancer, providing a real contribution to the mammography screening process.

Thesis Organisation

This thesis is organised into nine chapters. The first chapter is introductory, while the subsequent six chapters detail the development of the new CAD scheme. Final results are then presented followed by conclusions. As a consequence of this organisation, conclusions about specific CAD stages cannot be made until the final results chapter. For this reason several references forward to Chapter 8 are made.

Chapter 1 introduces the topic of breast cancer detection providing information on the New Zealand breast cancer screening process, survival statistics and a description of various cancerous diseases found in the breast. The section provides background knowledge of the current status in mammography screening and the limitations associated with this process. Reference is given to state-of-the-art computer aided detection schemes currently in the literature. The drawbacks of these schemes provide motivation for the development of the new CAD scheme detailed in the remaining chapters.

There are essentially two imaging modalities available in digital mammography: digitised film-screen mammograms and direct digital mammograms. Chapter 2 describes the different image characteristics of both modalities. This indicates that certain stages of a CAD scheme require separate development for each modality to ensure optimal detection results. Details of the mammogram databases used in the development and testing of the new detection scheme are also provided in Chapter 2.

The first stage of any potential CAD scheme involves identification of a region of interest to be processed in subsequent CAD stages. In this case the region of interest includes only breast tissue pixels. Chapter 2 details two segmentation methods for identifying the breast tissue and pectoral muscle regions of a mammogram. The multi-stage algorithm used in detection of the breast tissue is an original contribution in this thesis, while the algorithm for pectoral muscle detection is based on the straight line Hough transform.

Because microcalcifications are small high contrast features in mammogram images, their detection is limited by the presence of high spatial frequency noise. To accurately detect microcalcifications, measurement of the high frequency image noise is important. Chapter 3 details the development of a noise model for digitised film-screen mammograms based on a method previously developed in the literature. This modality has complex noise characteristics that strongly depend on the grey level. The noise model described is adaptive to varying imaging conditions, allowing the dependence of the noise on the grey level to be removed. An original in depth study into optimisation of the technique is carried out by investigating a number of alternative implementations aimed at improving the reliability of the noise estimates.

An original contribution is contained in Chapter 4 which details the development of a noise model for direct digital mammograms. It is found that quantum noise is the dominant noise source in this modality, ensuring a simple square root noise model can be applied. By processing experimental data taken at the University Medical Centre Nijmegen, Nijmegen, The Netherlands, an attempt is made to confirm the assumed quantum noise dominance. This is believed to be the first report to take full advantage of the linearity of a direct digital system in the detection of microcalcifications in direct digital mammograms.

The local operations used for the detection of microcalcifications often respond strongly to curvilinear structures (CLS). When detecting microcalcification clusters therefore, it is important that the appearance of both microcalcifications and CLS are modelled. Chapter 5 details the development of a shape parameter, based on the method of phase congruency. It is designed specifically to highlight CLS while simultaneously suppressing microcalcifications. Although the idea of phase congruency has already been established, the design of the shape parameter is all novel work.

Development of the overall microcalcification CAD system is detailed in Chapter 6. In the scheme a stochastic model is developed to enable the classification of individual pixels within a mammogram into separate classes based on Bayesian decision theory. The image features developed in Chapters 3 - 5 are used as input to this scheme. A principal advantage of the approach is that all the information available, i.e. the image data, the current labelling and prior beliefs, are exploited simultaneously. The CAD scheme is based on a well known method developed by Karssemeijer and is extended in this thesis to reduce the number of false positive detections related to CLS.

Chapter 7 details the process of grouping individual microcalcification objects into clusters. The presence of individual microcalcifications are in most cases not clinically significant. Therefore, it is important that a computer detection scheme distinguish between clustered and individual microcalcifications. The clustering scheme developed combines a mode clustering algorithm formerly proposed in the literature, with an original refinement algorithm based on Voronoi sets.

A detailed performance evaluation of the CAD scheme is provided in Chapter 8. This helps determine the accuracy of the system and hence the potential contribution to the mammography screening process. The aim is to determine the best system performance based on the various algorithm implementations detailed in previous chapters. Free-response receiver operating characteristic (FROC) analysis is used in determining the CAD performance.

Finally, conclusions and suggestions for future work are discussed in Chapter 9.

Supporting publications

The journal and conference papers resulting from the research described in this thesis are listed in chronological order below. The first conference paper is concerned with breast tissue segmentation detailed in Chapter 2; the next three conference papers cover the detection of microcalcifications when using the phase congruency shape parameter detailed in Chapter 5; and the final conference paper describes the hybrid clustering process discussed in Chapter 7. The journal paper, currently under peer review, covers the noise equalisation scheme for direct digital mammogram images described in Chapter 4.

K.J. McLoughlin and P.J. Bones, "Location of the breast-air boundary for a digital mammogram image," in *Proceedings of the Image & Vision Computing New Zealand conference (IVCNZ)*, Hamilton, New Zealand, November 2000, pp. 228-233.

K.J. McLoughlin, P.J. Bones and P.D. Kovesi, "Detection of microcalcifications in digital mammograms," in *Proceedings of the Image & Vision Computing New Zealand conference (IVCNZ)*, Dunedin, New Zealand, November 2001, pp. 259-264.

K.J. McLoughlin, P.J. Bones and P.D. Kovesi, "Connective tissue representation for detection of microcalcifications in digital mammograms," in *Proceedings of the SPIE Conference on Medical Imaging*, M. Sonka and J.M. Fitzpatrick, Eds., San Diego, California, February 2002, vol. 4686, pp. 1246-1256.

K.J. McLoughlin, P.J. Bones and P.D. Kovesi, "Detection of microcalcification clusters in digital mammogram images," in *Proceedings of the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 353-357.

P.J. Bones, K.J. McLoughlin and H.M. McLoughlin, "Data clustering for discrimination of microcalcifications in mammograms," in *Proceedings of the Image & Vision Computing New Zealand conference (IVCNZ)*, Auckland, New Zealand, November 2002, pp. 181-184.

K.J. McLoughlin, P.J. Bones and N. Karssemeijer, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images", *IEEE Transactions on Medical Imaging*, submitted July 2003, currently under peer review.

Abbreviations

CAD	computer aided detection
CC	cranio-caudal
CCD	charge coupled device
CLS	curvilinear structure
DCIS	ductal carcinoma in situ
DDA	digital differential analyser
DDSM	digital database for screening mammography
FDA	food and drug administration
FFDM	full field digital mammography
FN	false negative
FP	false positive
FROC	free response receiver operating characteristic
FSD-A	film-screen database A
FSD-B	film-screen database B
GE	general electric
ICM	iterated conditional modes
LCIS	lobular carcinoma in situ
MAP	maximum a posteriori
MIAS	mammographic image analysis society
ML	maximum likelihood
MLO	medio-lateral oblique

MPM	maximiser of posterior marginals
MRF	markov random field
MRI	magnetic resonance imaging
PPMA	polymethyl methacrylate
ROC	receiver operating characteristic
ROI	region of interest
RMS	root mean square
SA	simulated annealing
TDLU	terminal ductal lobular unit
TN	true negative
TP	true positive

Notation

$\approx$	approximately equal to
$\odot$	convolution operation
$\Im(.)$	imaginary part of
$\Re(.)$	real part of
∞	infinity
$\propto$	proportional to
$\sum$	summation
σ	standard deviation
μ	mean
$\|.\|$	magnitude
$O(.)$	order
$p(.)$	probability operator
$p(x\|y)$	probability of x given y
$E[.]$	population mean
$Var[.]$	population variance
$< ij >$	neighbourhood pair
$S\backslash i$	scene S excluding site i
$\underset{X}{\text{argmax}}(f(X))$	the value of X that maximises $f(X)$
$\max_X(f(X))$	the maximum value of $f(X)$ for all X
$\|.\|$	norm of a matrix or vector
$\in$	element of
δ_i	small neighbourhood of i
$\forall$	for all
$\neq$	not equal to

Chapter 1

Introduction

In many western countries breast cancer is the most common form of cancer among women. In New Zealand over 1500 women are diagnosed with the disease annually and about 600 women die from this type of cancer each year [18]. Over the past decades it has become apparent that breast cancer incidence rates are increasing steadily. Changes in risk factors such as diet and environmental factors seem to contribute to the rising incidence. Nevertheless, the mortality rates for breast cancer have remained relatively constant due to the development of more effective treatment and earlier diagnosis [45].

Many factors that increase the risk of the disease have been identified, such as family history of breast cancer, diet, late age of first birth, early onset of first menstruation and late age at menopause [17,105,128]. Despite this, these risk factors still can not explain the major part of the incidence. Since these factors can not be used to prevent the disease, the most important strategy to reduce breast cancer mortality seems to be early detection through organised breast cancer screening programmes. Early detection and diagnosis enables effective treatment of breast cancer and can increase the chance of survival for patients by up to 80%.

Since 1992 pilot projects for breast cancer screening have been carried out in New Zealand. Currently, the nation-wide breast cancer screening network, which began in 1998, is offered to women aged 50-64. Women in this age group are requested once every 2 years to have a breast examination using mammography (i.e. a technique for making images of the breast using x-rays). Limiting the age of women participating in the screening program not only ensures the health risk due to radiation is reduced as far as possible, it also keeps the program cost-effective in terms of costs per life year gained [18, 61, pp. 1-10, 138]. Mammography is widely regarded as the most effective method for detection of subtle abnormalities in the breast [85, pp. 1-27]. The mammogram can detect small changes in

breast tissue which may indicate cancers that are too small to be felt either by the woman herself or by a doctor. Other modalities have been applied to breast imaging including ultrasound and magnetic resonance imaging (MRI), however both these techniques are unable to reliably detect the small subtle signs of cancer [41, 55].

Screening for breast cancer is a complex task, due to the small fraction of malignant cases: approximately 5 out of 1000 women in the screening population have breast cancer [30]. With nationwide screening programmes in many western countries, the number of mammograms to be analysed by radiologists is enormous. Manual reading is both labour intensive and time consuming, demanding great concentration. Furthermore, because of the large number of normal patients in the screening programmes, there is a risk that radiologists may miss some subtle abnormalities. The fundamental conflict of a boring task with serious consequences of a mistake is something to which humans tend to be unsuited. The rapid advancement of computer technology, pattern recognition, digital image processing, and artificial intelligence has led to a new direction in mammogram analysis. To help radiologists in their task to detect signs of cancer a number of research groups are developing methods for computer aided detection (CAD). Radiologists can use prompts from CAD findings to improve their sensitivity (number of true positive detected cancers) and diagnostic performance.

When a cancerous process in the breast is at a very early stage, clusters of microcalcifications may appear as the only sign. Microcalcifications are small calcium deposits within the breast. On a mammogram they appear as regions of elevated intensity against the varying background density of the x-ray image [61, pp. 191-203]. However, 80% of all microcalcification clusters that are encountered are due to benign processes. Differentiation of (obvious) benign microcalcification clusters from malignant types during screening is regarded as essential. This is because recalling all cases with microcalcification clusters for further work-up (for instance, using magnification views and biopsies) would result in many false positives (cases that are recalled unnecessarily). This would cause a lot of unnecessary anxiety and thereby discourage women to participate. Furthermore, expenses of screening would increase considerably which is undesirable due to limited New Zealand health funds.

Characterisation of microcalcification clusters (i.e. differentiation between malignant and benign types) is known to be a very difficult task. Moreover, high intra and inter-observer variabilities have been reported. Accurate detection and segmentation of microcalcification clusters is an essential preliminary stage for a computer aided diagnosis scheme. The objective of the research described in this thesis was to develop a CAD scheme for the detection of microcalcification clusters with high sensitivity and low specificity (detection of false positive clusters). A distinction between benign and malignant microcalcification

clusters is not made therefore all regions suspected of containing a cluster are highlighted. The output from this detection scheme could be used as an initial stage for classification of clusters into benign or malignant types. It is not intended for the CAD scheme to replace the job of a radiologist, however it may be considered a useful tool in providing a second opinion. When a CAD scheme, such as the one presented in this thesis, can assist radiologists and thereby reduce the number of unnecessarily recalled women with microcalcification clusters, effectiveness of screening could be improved.

1.1 Breast diseases

In this section a description of the breast anatomy is given, followed by a short discussion on malignant and benign conditions that may appear in the female breast.

1.1.1 Breast anatomy

Histologically the breast is divided in 8 to 20 lobes or segments [61, pp. 1-10, 85, pp. 3-27, 157] (see Figure 1.1). Working backwards from the nipple, each lobe begins with a major duct that transfers milk to the nipple during lactation. This major duct branches several times, forming minor (or sub-segmental) ducts with correspondingly smaller diameters. Finally, the branching ducts end up in terminal ductal lobular unit's (TDLU's). A TDLU consists of a lobule and its extralobular terminal duct (final branch from the segmental duct). The lobule is responsible for milk production during the period of lactation and it is built up from 10 to 100 sac-like units called acini. The ductal and lobular system as a whole is surrounded by an uninterrupted basement membrane on the outside.

With age the breast tissue will change. In young women the breast tissue is dense and rich of glandular tissue. On aging, the glandular tissue is gradually replaced by fat. This increased fat content of the breast in older women makes their mammograms relatively easier to diagnose. This is one of the major reasons that screening programmes have excluded women below a certain age (50 years in the New Zealand screening programme) [61, pp. 1-10, 138].

1.1.2 Malignant diseases

It is postulated that most carcinomas (malignant tumours) in the breast arise in the TDLU's either inside or just proximal to the lobule. The basement membrane plays a key role in determining whether a tumour is *in situ* (has not grown through the basement membrane) or *invasive* (has grown through the basement membrane). When in situ carcinomas develop into invasive cancers, they can form metastases to lymph nodes and other organs which will decrease the survival chance. In the breast two forms of in situ malignancy can be con-

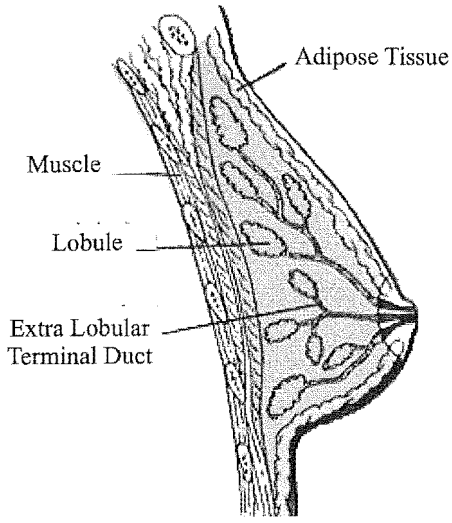


Figure 1.1 Anatomy of the breast (source: www.rmccares.org)

sidered: the ductal carcinoma in situ (DCIS) and the lobular carcinoma in situ (LCIS). DCIS can involve a variable number of ducts and lobules and often involves central necrosis (i.e. dead cells in the centre of the duct), but is always contained within the ductal or lobular system. DCIS is considered to be the earliest detectable stage of breast cancer. The term DCIS describes a heterogeneous group of lesions [61, pp. 1-10, 85, pp. 3-27, 157]. One can distinguish different patterns of DCIS including well differentiated, intermediately differentiated and poorly differentiated, where the later is the most aggressive. In most ductal carcinomas in situ microcalcifications can be found, which may result either from active cellular secretion or from calcified necrosis.

The lobular carcinoma in situ (LCIS) rarely produces microcalcifications and the diagnosis is usually an incidental microscopic finding due to other breast disorders leading to excision of breast tissue. When histologically only LCIS is found no further treatment will be performed. Although LCIS itself is not a risk of life it is regarded as a risk factor for developing well differentiated DCIS and invasive cancer [85, pp. 3-27].

1.1.3 Benign diseases

The following subsection discusses important benign conditions that show microcalcifications or morphologic growth patterns similar to carcinomas. These conditions are difficult to distinguish from a malignant process. One can consider a number of benign diseases that occur within the lobular and ductal system and that are associated with microcalcification clusters [62, 63, 85, pp. 3-27, 99, 135, 140].

1. Blunt duct adenosis may be characterised by calcium deposits that form calcifications within the sac-like elements of the lobules (the acini). Typically these are small punctate calcifications which appear uniformly faint or moderately intense on mammograms. The calcifications are separated by fine lines and they are usually collected in round or oval clusters.
2. Sclerosing adenosis occurs when an increase of fibrous tissue causes compression and deformation of lobular structures. The shape of microcalcifications which may occur in the acini, reflects the lobular distortion: they show considerable polymorphism compared to the uniformly punctate calcifications of blunt duct adenosis.
3. Microcystic adenosis, also known as cystic hyperplasia, may appear as elongated calcifications resembling the shape of a teacup, seen in oblique mammograms (side view). This effect is caused by milk or calcium that may settle in the dilated acini.

Other important benign conditions occur outside the ductal and lobular system [62,63,135, 140].

1. Arterial calcifications of the breast are seen as parallel calcifications of variable lengths.
2. Fibroadenomas are benign tumours developing from an over-growth of lobular connective tissue. Fibroadenomas are usually round, oval or lobulated, and can contain coarse and dense microcalcifications.
3. Fat necrosis may occur after trauma or inflammation. When fat cells are injured there is a release of fatty acids from the cells. The result is a fibrotic capsule around the oily substance. Calcifications can occur within this capsule and the central oily substance.

1.2 Mammography

Mammography is a diagnostic breast imaging method using x-rays. It is widely used to detect and characterise breast cancer and because of its high performance and low cost it is by far the imaging technique most suited for screening programmes.

1.2.1 Mammographic abnormalities

Microcalcification clusters not only appear in both in situ and invasive breast cancer but also in benign diseases. Many of the breast cancers that are at an early stage are currently detected by the presence of microcalcifications. Only when appearing in a cluster of three or more calcifications are they considered clinically suspicious. Microcalcifications that are visible in mammograms vary in diameter roughly from 100μm to 500μm. Figure 1.2 shows four clusters. The clusters on the right ((b) and (d)) are benign whereas those presented on

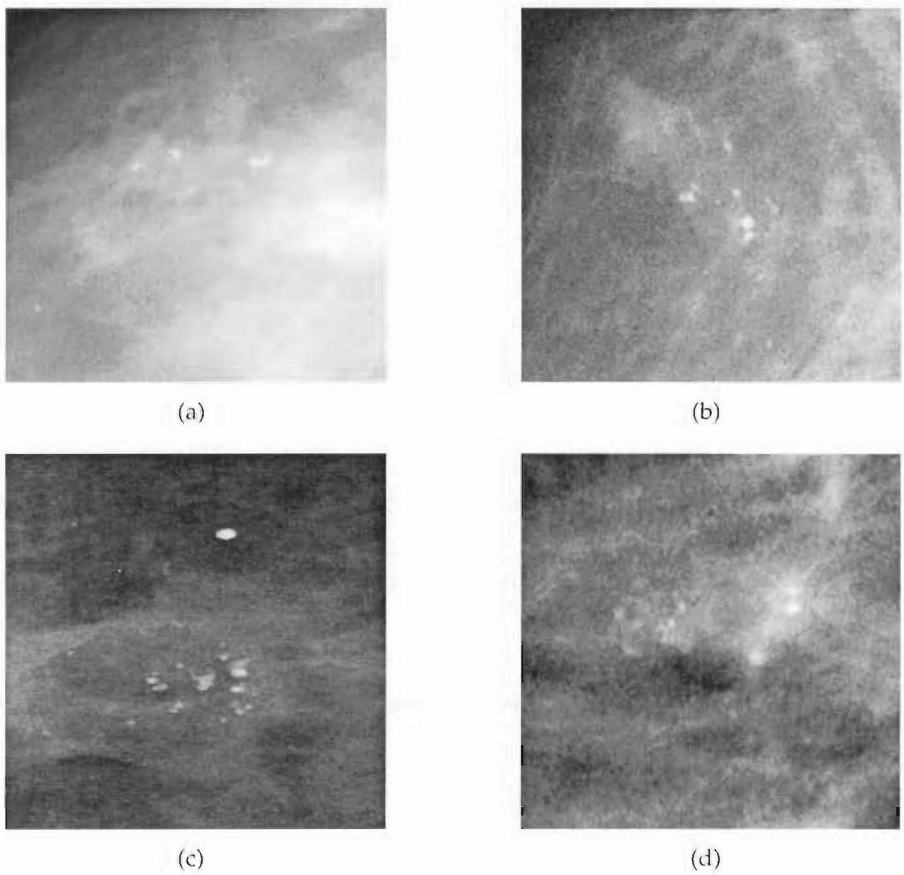
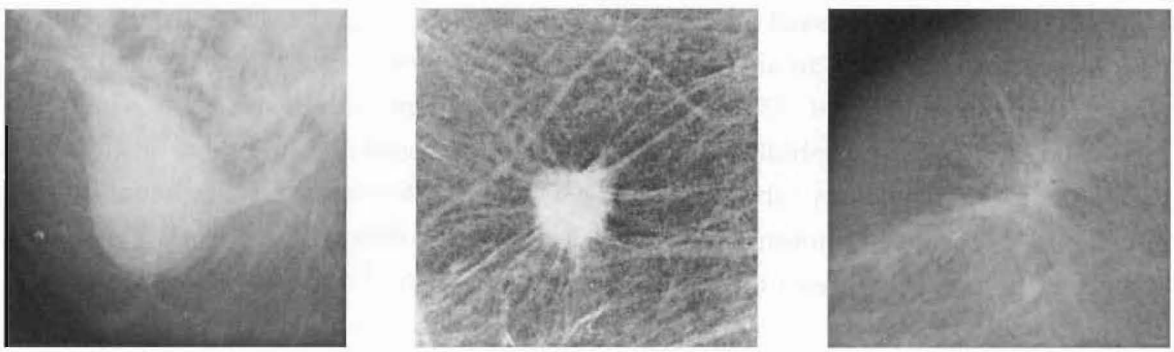


Figure 1.2 Benign (left) and malignant (right) microcalcification clusters.

the left ((a) and (c)) are malignant. Differentiation between malignant and benign clusters, based on mammographic appearance, is not an easy task.

Apart from microcalcification clusters, one can classify the visual signs for which radiologists search during mammographic screening into three basic categories: masses, architectural distortions and asymmetric densities [56]. These abnormalities may indicate invasive breast cancer. Masses that are sharply defined (circumscribed masses) are usually benign. However, if a mass has a faint jagged edge it is likely to be malignant. A mass surrounded by a radiating pattern of spicules is called a spiculated mass or stellate lesion. Stellate lesions are highly suspicious indicators of breast cancer.

Not all tumours have a central mass, some carcinomas are often only detectable due to an architectural distortion of the breast tissue. These lesions are often quite subtle and can occur with both benign and malignant processes. Figure 1.3 shows typical examples of a mass, stellate lesion and an architectural distortion.



(a) (b) (c)

Figure 1.3 Examples of most common signs of malignant abnormalities: (a) circumscribed lesion, (b) stellate lesion, and (c) architectural distortion.

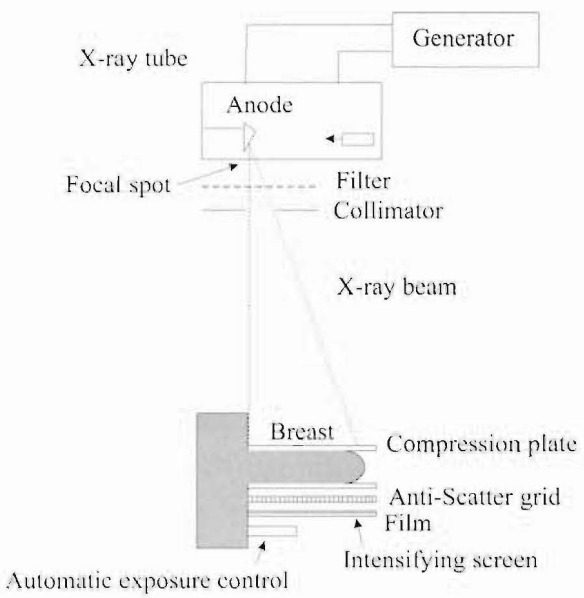


Figure 1.4 Schematic representation of the components of a conventional film-screen mammographic system.

Some masses are detected by radiologists because of asymmetry in the breast pattern between the left and right breast. Asymmetry may be a suspicious sign because in a normal breast the fibro-glandular breast pattern is often symmetric with respect to both breasts. To increase the sensitivity and specificity, mammograms from previous screening rounds are used to detect changes between the old and new films.

1.2.2 Conventional and digital mammography

Figure 1.4 shows a schematic representation of the components of a conventional film-screen mammographic system. To obtain a mammogram image a beam of x-ray photons is

directed towards a compressed breast. The x-rays form a divergent beam from the source, whose focal spot and anode are situated in the x-ray tube above the chest wall of the woman being examined [31, 85, pp. 135-155]. The incident x-ray beam is polyenergetic, that is, it comprises x-ray photons of many different energies and has a spectrum which varies according to the x-ray tube voltage, anode (target) material and subsequent filtering. The energy of the x-ray photons is important because the difference between the attenuation coefficients of the breast tissues rises with lower energy and thus there is a greater difference in the x-ray signal exiting the breast when a lower energy beam is used. This, in turn, leads to more contrast in the image. However, lower energy photons are more readily attenuated in the breast tissue leading to a greater radiation dose. Thus there is a trade-off between contrast in the mammogram and radiation dose to the breast. In order to remove the low energy photons before they reach the breast, a filter is placed directly below the anode. Regulation of the size and shape of the x-ray beam is performed using a collimator, which limits the amount of superfluous exposure, or scatter reaching the breast. As the x-ray photons travel through the breast they are attenuated by both scattering and absorption processes. The probability of attenuation is related to the type of tissue through which the photons have to pass. When the x-ray signal exits the breast it has components due to photons passing straight through the breast (primary radiation), photons that have been scattered, and photons that come from other than the focal spot (extra-focal radiation). It is the primary radiation that contains the most useful information for diagnosis. In order to reduce the other two components an anti-scatter grid is positioned between the breast and the film-screen combination. The use of an anti-scatter grid however requires the radiation dose to be higher. X-ray photons passing through the breast and anti-scatter grid also have to pass through the film before being absorbed by an intensifying screen which produces visible light. The light given off by the intensifying screen exposes the film and creates the image. Exposure to the breast is stopped once an automatic exposure control, positioned under a section of the breast, has received a set exposure.

Abnormalities in the breast tissue are often very subtle. Therefore, the mammography machines, film, and developing process are specially designed to create mammograms that are sensitive for these subtle differences. In order to get a good image the breast is compressed, which also enables lowering the dose of radiation. In a standard examination, two images of each breast are taken: one from the top (called a cranio-caudal or CC view) and one with the x-ray tube angled approximately 45° medially (called the medio-lateral oblique or MLO view). This ensures that the images display as much breast tissue as possible.

Although conventional (or film-screen) mammography has high sensitivity and specificity, it has some important limitations as well [85, pp. 135-155, 157]. The film used to capture, store and display the mammographic image is one of the major technical restrictions of film-screen mammography. The visibility of breast cancer depends on different attenuation of the x-ray beam by the suspect regions compared with the surrounding tissue. Sus-

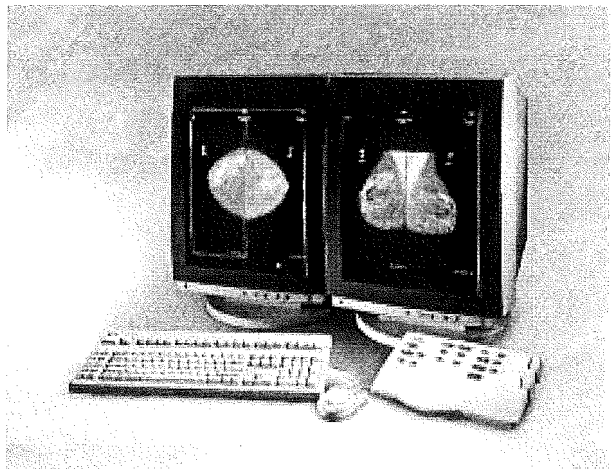


Figure 1.5 Mammographic images displayed on two monitors.

pect regions lying in dense areas of the breast may not be noticeable because film contrast decreases in the densest breast areas. This is due to the limited dynamic range of conventional films. Furthermore, the image data obtained using a film-screen system cannot be manipulated once the image is processed in a film processor. Specifically, over and underexposed images have to be recorded again. Contrast levels in the image cannot be altered to improve the relative visibility of structures in the image without recording additional images of the patient.

Digital mammography has advantages over conventional mammography [44, 59, 85, pp. 135-155, 157]. Image acquisition and display are decoupled in a direct digital imaging system (also known as Full Field Digital Mammography (FFDM)). A digital detector provides an electronic signal proportional to the intensity of the x-rays transmitted through the breast. When digitised and stored in computer memory, the image can be processed by a computer, and either printed on film or displayed on a monitor (Figure 1.5). The steps in the breast imaging chain can each be optimised since they are separated. Storage, transmission and retrieval of images can be improved also. Computer aided diagnosis software can be applied to help the radiologist interpret the images. Furthermore, digital detectors have higher dynamic range, and may require a lower x-ray dose [96].

There are different kinds of direct digital mammography detectors currently available clinically. For instance, full-field detectors are developed that consist of an array of light sensitive diodes deposited on a plate of amorphous silicon and covered with a cesium iodide x-ray absorbing phosphor [40]. Other systems use a phosphor screen combined with fiber-optic tapers that transfer the light to charge coupled device (CCD) arrays [97,157]. Current systems record mammographic images at spatial resolutions of 50μm to 100μm per pixel.

Since four mammographic images (two of each breast) cannot be viewed at full spatial resolution simultaneously on existing monitors on the market today, digital mammograms can be printed to film as an alternative. However, using a film display loses the potential benefits of digital mammography in terms of interactive manipulation of the image. More practical displays will be needed in order for the full potential of digital mammography to be reached

1.3 Computer aided detection

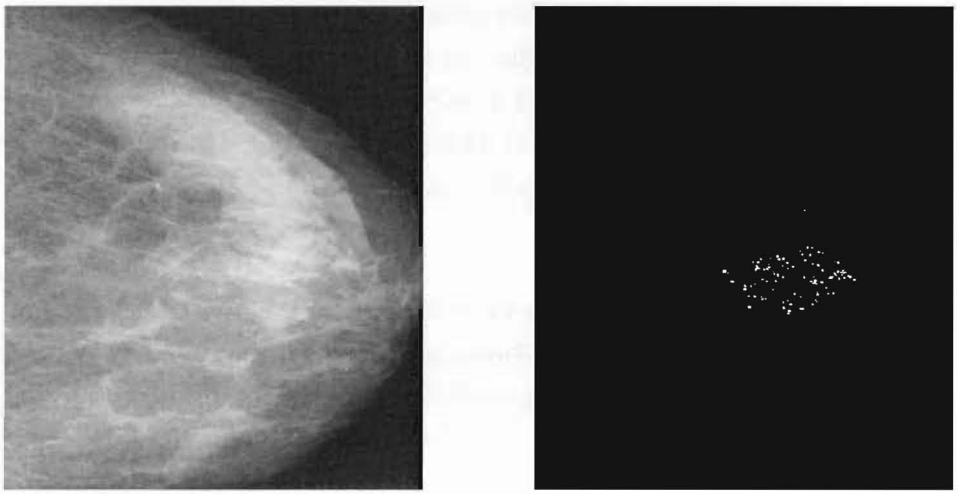
Most of the research in computer assisted reading of mammograms focuses on developing methods for detection of abnormalities, like microcalcification clusters, densities and stellate lesions [76]. Successful results are reported in a number of studies [6,8,19,24,47,50,76, 81,151]. These studies demonstrate a significant increase in radiologist screening efficacy when using computer aided detection (CAD).

This section outlines specific tasks CAD can be used for, and gives references to various important approaches and algorithms.

1.3.1 Microcalcification clusters

A variety of microcalcification detection methods have been reported in the literature. Most methods have in common that one or more filters are used to determine the local contrast at each pixel inside a region of interest, usually representing the whole breast. Such a local contrast measure is determined by high frequency image noise and local contrast structures occurring in the mammographic image. Detection of microcalcifications is basically searching for high local contrast features, caused by the high x-ray attenuation of microcalcifications compared to their background.

It has been shown that it is not the spatial resolution but the high spatial frequency noise that limits the detectability of microcalcifications [28, 29, 75]. In order to detect microcalcifications, therefore, reliable measurement of the high frequency image noise is of crucial importance. Most microcalcification detection methods described in the literature use some noise dependent adaptive threshold that is locally determined. This is important for dealing with variation of the noise level across the image [24, 114, 117]. Other researchers avoid using locally adaptive thresholds by taking the signal dependency of noise into account. In their detection schemes, a pre-processing step of noise equalisation is performed which makes the noise level independent of the grey level. Using such an approach in methods for microcalcification detection is found to be beneficial in a number of studies [71, 98, 115, 124, 150]. Additional approaches are based on adaptive noise suppression and other hybrid wavelet adaptive signal enhancement methods [125, 126, 141].



(a) (b)

Figure 1.6 Automated detection of microcalcifications. (a) Cranio-caudal view of a breast with a microcalcification cluster. (b) Output of an automated detection method, where the white spots represent the microcalcifications detected.

Apart from microcalcifications, other high-contrast structures exist that have high local contrast, such as vessel walls and thin strings of connective tissue. Furthermore, peaks in the high frequency image noise may be hard to distinguish from microcalcifications. Simply selecting pixels with high local contrast yields many false positive findings. Microcalcification detection schemes based on an initial detection step followed by a second step of classification that is intended to remove false positive findings were developed by many researchers. Common approaches that have been used include rule-based methods [113, 132] and artificial neural networks [48, 158]. With respect to performance, the current state of the art is that the sensitivity of methods for detecting microcalcification clusters can be very high, up to 90% at a false positive rate of approximately 1 per image. Figure 1.6(a) shows a CC (cranio-caudal) mammogram with microcalcifications. The output image (Figure 1.6(b)) shows a cluster of microcalcifications that was detected by the statistical detection method described in [71].

Few studies are related to computer aided classification (differentiation) of benign and malignant microcalcification clusters. Features described in the literature for the classification schemes include distribution features (based on the distribution of microcalcification properties within a cluster), cluster shape features and texture features. The latter type of features describe texture properties of the local surrounding tissue. Most researchers use either human extracted features [10, 156], or manual detection of calcifications along with computer extracted features as the input for their automatic classification system [27, 67]. Automated classification applied to microcalcifications that are manually detected by radiologists avoids problems related to false positive or false negative detected microcalci-

fications that occur when using automated detection. However, such an approach hampers clinical use of automated interpretation of microcalcification clusters. Schmidt *et al* addressed this draw-back and developed a fully automated detection and classification method for microcalcification clusters [134]. However, the author found the computer system could not infer a reliable diagnosis with respect to cases rated by human experts as being hard to diagnose.

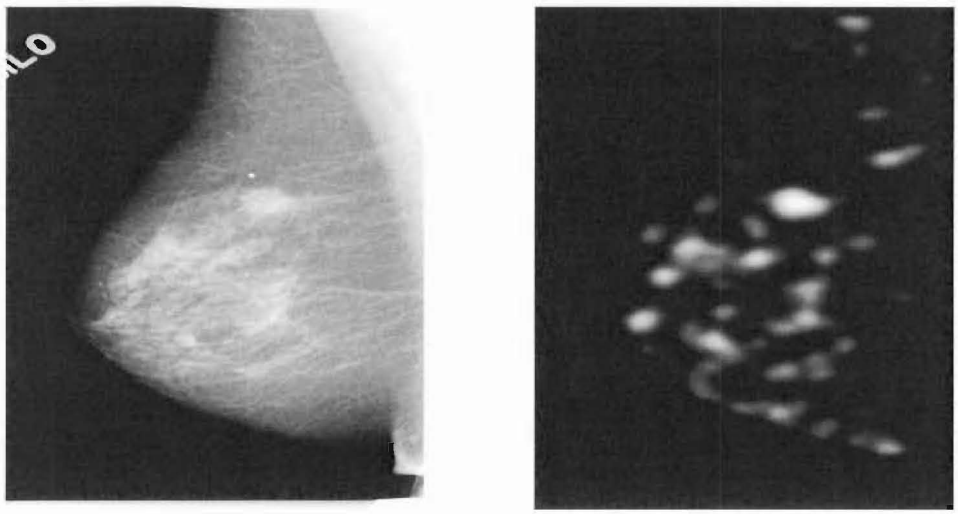
Results reported in the literature so far are promising. For instance Jiang *et al* compared his method with the mean performance of radiologists in two studies using two different data sets. In both studies his scheme outperformed the radiologists significantly [67]. It must be noted that this method is not fully automated since microcalcifications are manually identified.

1.3.2 Masses, architectural distortions and asymmetric densities

A large number of segmentation methods have been developed in the field of image analysis and many of them have been used to segment masses in mammograms. One of the most popular segmentation methods, used in many image processing fields, is region growing. Region growing has been applied to the computer aided detection of mammographic lesions by a number of groups [87, 93, 107, 123]. Another mass segmentation method developed by te Brake *et al* used a discrete contour model. The authors found that their method gave better segmentation results than a probabilistic region growing method [145].

The common approach for detecting lesions suspect for invasive cancer, is generation of target regions by spatial filtering, sometimes involving a measure of asymmetry between the left and right breast. In a second step the most suspicious regions are selected by a statistical or neural network classifier. Very sensitive algorithms have been developed for masses and stellate lesions, based on analysing statistical variations of local edge and line orientations [77,130,159]. Karssemeijer *et al* used gaussian second order directional derivative filters to locate spicules in a mammogram [77]. Zwiggelaar *et al* used a generic method of representing patterns of linear structures. This relies on the use of factor analysis to separate the systematic and random aspects of a class of patterns [159]. Sahiner *et al* applied an active contour model to detect a central mass, followed by radial gradient analysis to detect spicules located around the central mass [130]. Figure 1.7 shows an MLO (mediolateral oblique) mammogram and the corresponding output of a CAD system developed by Karssemeijer *et al*, where the largest bright spot is at the location of a histologically verified malignant mass [77].

Interestingly, good detection results have been reported on cases that were classified retrospectively as missed by screening [19, 144]. It must be noted that the sensitivity for de-



(a) (b)

Figure 1.7 A MLO image and the corresponding computer output of a mass that is automatically detected. The intensity level of the spots is a measure of the suspiciousness. The largest bright spot is at the location of a histologically verified malignant mass (Reproduced from [77]).

tection of masses, asymmetry and architectural distortions is still lower than for microcalcifications clusters. On the other hand, the number of false positive prompts generated by these methods is steadily decreasing, using the benefits of increased computational power of computers and the availability of huge databases of digital mammograms for training of algorithms.

Classification of benign and malignant masses is a well studied subject in mammography. All papers focus on edge analysis of the mass. A vague or spiculated edge indicates malignancy, whereas a sharp well defined contour is likely to belong to a benign abnormality [64, 123]. Additional features that are sometimes used are size, shape, texture and contrast measures. Interesting work was done by Huo *et al* [64] who used a radial edge gradient method to discriminate malignant and benign lesions, and Pohlman *et al* [123] who developed a feature describing tumour boundary roughness.

Chapter 2

Image characteristics, data and pre-processing

2.1 Image characteristics

An imaging system for the detection and diagnosis of breast cancer must provide visualisation of the key cancerous signs. Often these are very subtle features such as the fine detail associated with microcalcifications or the thin fibres radiating from a tumour mass. In order to separate the subtle changes in x-ray attenuation, the imaging system must precisely measure the transmitted x-ray intensity through all regions of the breast. Additionally, high resolution image acquisition and display is required to ensure detection of features as small as 100μm. Because of these requirements, mammography is one of the most technically demanding radiological imaging techniques [124].

Film-screen and direct digital modalities possess very different image characteristics, which affect, for example, the image contrast and patient dose required. Film has a complicated, sigmoidal-shaped response to the light the x-ray photons stimulate in the intensifying screen (see Figure 2.1(a)) [85, pp. 135-155, 157]. The best contrast separation is obtained using exposures that record the x-ray attenuation on the steeply sloped near-linear portion of the curve where small differences in tissue contrast are amplified. At low levels of exposure there is little change in density (the *toe* of the curve), producing little contrast between different tissues with high x-ray attenuation such as glandular tissue, masses and microcalcifications. Greater exposure is required to enhance the attenuation differences and provide contrast in the high attenuation areas of the breast. Similarly, at high exposures, the film density changes level out (the *heel* of the curve) and contrast separation is diminished. The non-linear response of film to x-ray exposure is a significant limitation of the film-screen modality.

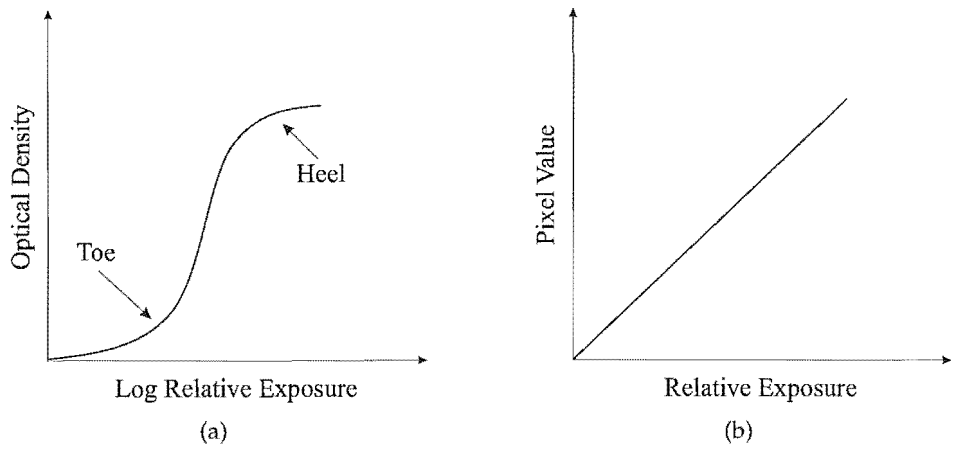


Figure 2.1 Typical characteristic curves for (a) mammographic film-screen combination, (b) direct digital mammogram.

Direct digital detectors have, essentially, a straight line response to x-ray photons, regardless of the level of exposure (see Figure 2.1(b)). This permits a wide dynamic range and maximum contrast separation, potentially allowing the limitations of film-screen mammography to be overcome [85, pp. 135-155].

Film-screen mammograms, such as that shown in Figure 1.7(a), invariably possess higher grey level values in the higher attenuation areas of the mammogram image (i.e. pixels associated with microcalcifications). The grey level of a pixel is defined to be within the range 0 (black) up to 2^b (white) for an b-bit image. Direct digital mammograms however are inverted, possessing higher grey levels in the lower attenuation areas of the image (i.e. background pixels). These images are known as *raw* direct digital mammograms, an example of which is shown in Figure 2.12(a). A digital imaging system is capable of processing the raw images to ensure they are visibly indistinguishable from film-screen mammograms. This processing is done before screening or diagnosis is performed by a radiologist. Throughout this thesis, all images referred to as *direct digital* are the raw images produced by a direct digital system.

2.2 Image databases

Currently there are three publicly available digitised film-screen databases used to facilitate research in the development of computer algorithms for mammography screening. The first is the Nijmegen database which consists of 40 digitised mammogram images obtained from the Nijmegen medical centre, the Netherlands [70]. All images are digitised using an Eikonix 12 bit CCD camera at a resolution of 100μm. This database is designed specifically for testing computerised microcalcification detection programs, with each image showing one or more clusters of microcalcifications. Annotations are provided with the database marking the position and size of all microcalcification clusters.

A second database known as the Mammographic Image Analysis Society (MIAS) database [143] contains 320 digitised film-screen mammograms. Each mammogram is digitised using a Joyce-Loebl microdensitometer, at 50μm resolution and 8 bits per pixel. Like the Nijmegen database it also includes radiologists annotations on the locations of any abnormalities (mass or microcalcification) that may be present.

The third and most extensive digitised film-screen database is the Digital Database for Screening Mammography (DDSM), produced at the University of South Florida [58]. This database contains 2620, four-view, mammography screening exams. Each study includes two images of each breast, along with associated patient information (i.e. age at time of study, subtlety rating for abnormalities, keyword description of abnormalities) and image information (i.e. scanner, spatial resolution). Images containing suspicious areas (either benign or malignant) have associated *overlay* files which provide annotations marking the location and size of each abnormality.

Since the DDSM is the most recent and extensive digitised film-screen database available, images from this database have been selected for development of the CAD scheme described herein. It should be noted that collection of images in this database was subject to the Institutional Review Board (IRB) process [58]. The first data set chosen from the DDSM consists of 50 arbitrarily chosen mammogram images, composed of both mediolateral oblique (MLO) and cranio-caudal (CC) views from 25 different patients. All mammogram images contain malignant abnormalities (either mass or microcalcification), the location and pathology of which is provided by the overlay files associated with each mammogram. The images are film-screen digitised at a resolution of 50μm and 8 bits per pixel, using various digitisation systems. To avoid extra computer processing each mammogram is cropped, removing any non-exposed area in the film-screen image. Appendix A.1 provides a list of the exact images used for this data set, which are referred to as Film-Screen Dataset A (FSD-A).

A second data set chosen from the DDSM is designed specifically for the training and testing of a microcalcification CAD scheme. The data set consists of 50 mammogram images (different from FSD-A), composed of both MLO and CC views from 32 different patients. A total of 37 malignant microcalcification clusters are present. Each image is digitised at a resolution of 50μm and 8 bits per pixel, using various digitisation systems. Once again to avoid extra computer processing each mammogram is cropped to remove any non-exposed area in the film-screen image. Appendix A.2 provides a list of the images used for this data set, which are referred to as Film-Screen Dataset B (FSD-B).

The direct digital modality is a very recent development in the area of x-ray mammography, therefore no publicly available database has been created for the purpose of computerised mammography screening. A set of direct digital mammogram images were ob-

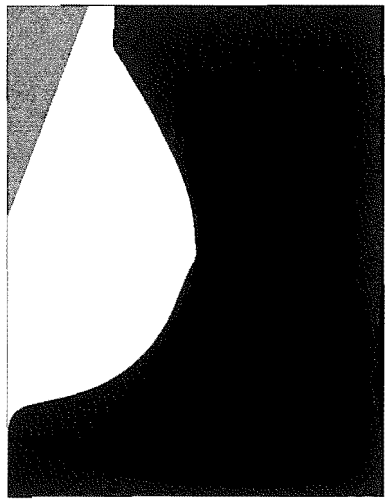


Figure 2.2 Identifying three distinct areas within a mammogram image: breast tissue (white), pectoral muscle (grey) and background (black).

tained from the General Electric (GE) Senographe system at the Nijmegen medical centre. Note that in The Netherlands, informed consent is not required for the use of anonymous patient data for scientific purposes. The patient's name was not identified in any of the direct digital images used here. The data set consists of 124 mammograms composed of both MLO and CC views from 33 patients. Twenty eight of the mammograms contain a single microcalcification cluster (14 benign and 14 malignant). Labelling of the microcalcification clusters was performed by an expert radiologist who provided both the pathology and location of each cluster. All mammograms are 2294×1914 pixels in size and have a 100μm resolution. The original 14 bits per pixel images were linearly rescaled to 12 bits (a standard precision used in medical imaging).

The data sets described above are used for the testing and training of all microcalcification CAD stages described herein.

2.3 Pre-processing

Extraction of quantitative features for classification of mammograms requires accurate and robust segmentation methods for separating the breast tissue and pectoral muscle from the background (see Figure 2.2). More than one-third of a mammogram is background containing very little useful information to a CAD system [95]. Removing the background, therefore, significantly reduces the processing time and improves the efficiency of subsequent computer programming stages. Automatic determination of the pectoral muscle is useful in restricting the search space for cancer detection, determining mammogram adequacy as well as mammogram-pair registration and comparison [88].

The remainder of this section provides details of a new, fully automated method for subdividing the mammogram into three distinct areas: breast tissue, pectoral muscle and background. Subsection 2.3.1 gives a review of the breast tissue and pectoral muscle segmentation techniques previously developed in the literature. Subsection 2.3.2 describes the format of the mammogram images used for training the new segmentation scheme. The steps involved in identifying the breast air boundary are described in detail in subsection 2.3.3. The algorithm utilises methods of thresholding and active contour modelling in a unique way, thus obtaining an accurate breast outline. Finally, subsection 2.3.4 describes a method for segmenting the pectoral muscle using a modified straight line Hough transform [42,73]. It should be noted that development of the algorithms in this chapter are carried out using digitised film-screen mammograms and adjustments are mentioned for application to direct digital mammograms.

2.3.1 Existing methods

Many authors have developed methods to identify the breast-air boundary in digitised film-screen mammograms. Both global histogram analysis and morphological filtering techniques have been used in a number of these studies [51, 52, 158]. Although these techniques have the advantage of simplicity they are however extremely sensitive to the parameters chosen. For example, a method based on thresholding alone is critically dependent on the selection of the threshold value. Also, the size and shape of the structuring element in morphological operations can significantly effect the fidelity of the extracted breast boundary when compared to the original.

Other researchers have incorporated several stages into their breast boundary detection algorithms to eliminate the above mentioned problems. Mendez *et al* uses a border tracking method based on the gradient of image intensity [106]. First, the grey level histogram of the digitised mammogram is calculated. Two cut-off grey levels are obtained from this histogram to segment the unexposed and directly exposed non-breast regions. To avoid the influence that small details have on the final result the mammogram image is smoothed, eliminating the presence of local maxima. Finally, a simple tracking algorithm is applied perpendicular to the expected breast border. A point (x, y) belongs to the border if the grey level value $f(x_i, y_i)$ of the nine previous pixels verify the condition

$$f(x_1, y_1) \leq f(x_2, y_2) \leq \cdots \leq f(x_7, y_7) \leq f(x_8, y_8) \leq f(x_9, y_9) \leq f(x, y). \qquad (2.1)$$

Results obtained when applying this method identified several drawbacks. Often the segmentation produces a breast boundary inside the manually identified outline. This discrepancy arises due to the algorithm's dependence on a steady grey level gradient at the breast edge, which is not guaranteed. Also, no attempt is made to provide a smooth contour as

would be expected on the breast boundary, and the resulting outline is often irregular and discontinuous.

One of the most well known breast boundary detection algorithms, developed by Bick *et al*, determines a binary breast mask based on modified histogram analysis after median filtering [14]. Median filtering is used as an initial stage to reduce noise and eliminate isolated pixel artifacts. A selective modified histogram is constructed including only pixels with a small local grey level variance. Pixels are then classified, using the modified histogram, as either direct-exposure, unexposed or potential object. Morphological filtering eliminates minor irregularities along the breast contour which is subsequently smoothed using a running average of the border coordinates. This method was tested on a large film-screen image database using various digitisation systems. Minor discrepancies were found in some outlines due to the fact that smoothing of the contour does not utilise the image characteristics of the mammogram (i.e. greyscale or image gradient).

With respect to pectoral muscle segmentation, the straight line Hough transform [42, 73] has been regarded as the most robust and generally acceptable technique. A detailed description of this method is given in Section 2.3.4, including additional constraints to optimise the transform for pectoral muscle detection.

2.3.2 Format

The following segmentation techniques are performed on a low-resolution copy of the mammogram, to reduce processing time and to avoid the influence that small details have on the final result. Each mammogram is downsampled using a polyphase implementation, producing a 250μm low-resolution image. In this description it is further assumed that the image is positioned with the chest side on the left of the digital image matrix. A coordinate frame is used in which the origin is at the centre of the image matrix, positive x corresponds to the right and positive y to the upper part of the image.

2.3.3 Breast air boundary

Detection of the breast border should be the first step in any image computerised analysis. CAD methods require an initial stage of processing to locate and mask the breast region, removing the off-breast noise field and eliminating patient markings that are often of very high intensity. There is an obvious need for an algorithm that is not extremely sensitive to parameter selection, and produces an accurate and smooth outline of the breast region. Figure 2.3 provides an overview of the image processing steps required for a new fully automated breast-air segmentation scheme. Details of the algorithms are to follow.

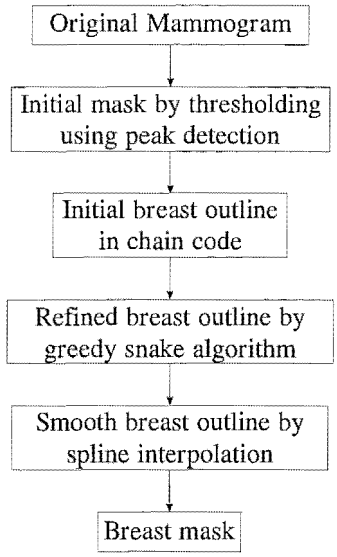


Figure 2.3 Sequence of steps in the new breast air segmentation scheme for film-screen digitised mammograms.

Thresholding using peak detection

Initial segmentation of the breast tissue area requires application of a simple global thresholding technique, separating the majority of the breast from the background. An extensive survey of thresholding methods is provided by Sahoo *et al* [131], however many of these techniques are not applicable to breast border segmentation due to the unique properties of the mammogram image. The background in a mammogram has a very uniform low intensity, which shows up as a dominant peak in the grey level histogram. Even the most well known thresholding schemes such as Otsu [120], moment-preserving [148] and Lloyd-max quantisation [94] produce threshold values too high for application to breast border detection. For this reason a peak detection method is implemented in preference to the standard global thresholding techniques.

The peak detection method proposed by Sezan automatically determines a threshold from the grey level histogram of the mammogram [136]. This detection scheme generates a peak detection signal r_N by convolving the grey level histogram $h(f)$ with a peak detection kernel p_N

$$r_N = p_N \odot h(f). \tag{2.2}$$

The peak detection kernel is given by

$$p_N(n) = \begin{cases} -\left(\frac{N-1}{2} + n + 1\right) \cdot \frac{1}{N}, & -\frac{N-1}{2} \leq n \leq -1 \\ \left(\frac{N-1}{2}\right) \cdot \frac{1}{N}, & n = 0 \\ \left(\frac{N-1}{2} - n\right) \cdot \frac{1}{N} & 1 \leq n \leq \frac{N-1}{2} - 1, \end{cases} \tag{2.3}$$

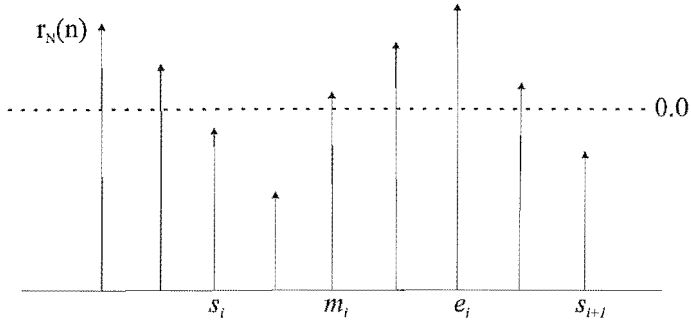


Figure 2.4 Illustration of an ideal detection signal and the parameters that characterise the peaks.

where N is the detection parameter (assumed to be odd). This parameter controls the sensitivity of the peak detection and hence, the number of detected peaks. As N decreases the result of the convolution becomes more sensitive to local variations such as single peaks and possibly noise. For larger N, the algorithm is more responsive to global variations.

To estimate the start, end, and maximum points of the peaks in the image histogram the following principles are applied to the detection signal r_N:

1. A zero-crossing of the detection signal to negative values (negative crossover) indicates the start of a peak. The grey level at which the negative crossover occurs is defined to be the estimate of a start point. For the ith peak, this grey level is denoted by s_i. Similarly, the next negative crossover at the grey level s_{i+1} estimates the start of the next peak.
2. A zero-crossing of the detection signal to positive values (positive crossover) following a negative crossover estimates the grey level at which the peak attains its maximum. This grey level is denoted by m_i.
3. The grey level between two successive negative crossovers at which the detection signal attains its local maximum is defined to be the estimate of the end point of the peak in the image histogram. For the ith peak, this grey level is denoted by e_i.

In general, a peak is represented by the triplet: (s_i, m_i, e_i). The parameters s_i, m_i, and e_i are illustrated in Figure 2.4 in the case of an ideal detection signal. Determining the global threshold for breast tissue segmentation requires the end of the low intensity background peak in the image histogram to be found. Therefore, e_1 is chosen as the global threshold providing initial segmentation of the background from the breast tissue.

Figures 2.5(c) and (d) show both the grey level histogram, $h(f)$, and peak detection signal, r_N, generated for the film-screen mammogram in Figure 2.5(a). The peak detection signal

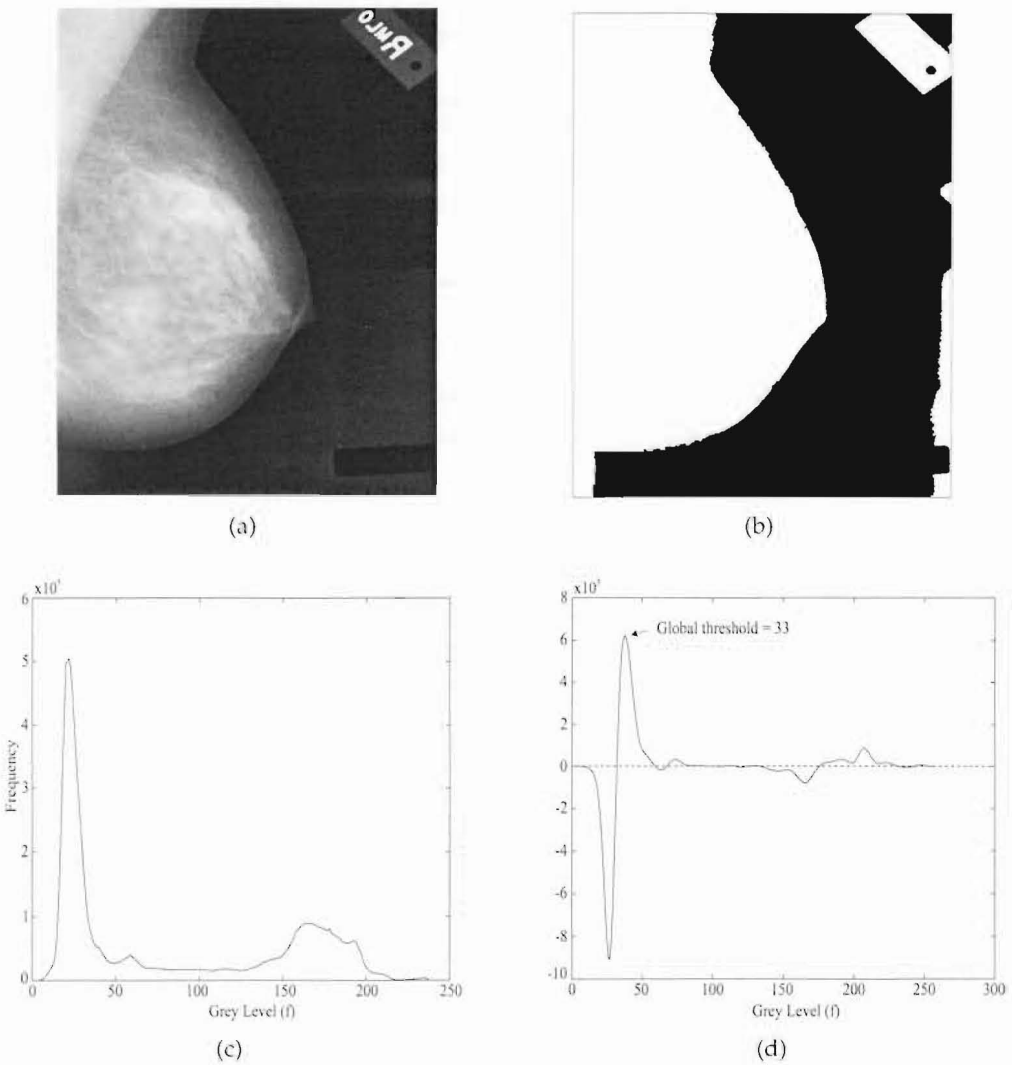


Figure 2.5 (a) Original film-screen MLO mammogram image. (b) Binary breast mask using a global grey level threshold of 33. (c) Grey level histogram, $h(f)$. (d) Peak detection signal, r_N, for $N = 17$.

was obtained using a detection parameter $N = 17$. Also indicated in Figure 2.5(d) is the resulting global grey level threshold of 33. Applying this threshold to the original mammogram produces a binary mask of the breast tissue. Figure 2.5(b) shows the masked breast tissue after removal of isolated pixels using morphological opening (erosion followed by dilation). Both the breast tissue and the film labelling have been segmented, however the breast tissue is easily identified by searching for the largest binary object within the mask.

Table 2.1 provides threshold values obtained when applying various other global thresholding techniques to the mammogram in Figure 2.5(a). All techniques produce thresholds considerably higher than that of the peak detection method, generating binary masks that do not cover the entire breast tissue area. This reinforces the fact that the peak detection method is more suited to this particular application.

Table 2.1 Global threshold values obtained when applying different thresholding techniques to the mammogram image in Figure 2.5(a).

Threshold Technique	Threshold Value
Lloyd-max	96.5
Otsu	96
Moment-preserving	103

It is important in any algorithm development to have limited sensitivity to changes in parameters. The sensitivity of the peak detection scheme was determined by applying a range of detection parameters, N, to FSD-A. The optimal detection parameter for each mammogram is dependent on the width of the background low intensity peak in the corresponding image histogram. It was found however, that the global threshold obtained from the algorithm was insensitive to large changes in N (acceptable range $N = 11$ to $N = 25$). This is due to the dominance of the background peak in the grey level histogram, thus indicating that the method is relatively robust.

Boundary representation

Finding the edge pixels of the binary mask allows further manipulation of the segmented boundary to provide a smoother, more accurate representation of the breast area. A chain code is used to represent the boundary by a connected sequence of straight line segments [54, pp. 392-394]. Eight-connectivity of the pixels is used and the direction of each segment is coded using a numbering scheme. The start point is the first pixel found within the breast tissue region when scanning the mask image from left to right and top to bottom. The chain code traverses clockwise around the boundary of the binary mask, assigning a direction to the segments connecting each pair of boundary pixels. Figure 2.6 provides an illustration of the boundary found for the binary mask in Figure 2.5(b).

Greedy snake algorithm

The breast boundary obtained after chain coding is often jagged and possesses small deviations from the actual breast contour. An algorithm is required to refine this initial outline, providing a smoother boundary more accurately aligned with the breast edge. Active contour models (or snakes), originally proposed by Kass *et al* [79], have been used in many applications for locating boundaries and features of interest in images [69,90,91,127,130,146]. Simply defined, a snake is a deformable contour that moves under a variety of local image and object-model constraints. The snake is controlled by minimising a function which converts contour information (i.e. length and curvature) and image information (i.e. edge gradients) into energies making it very suitable to the application of breast boundary refinement.

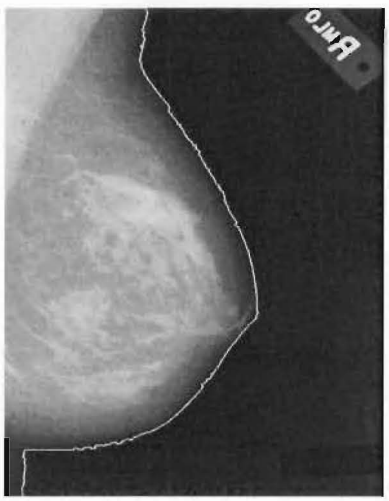


Figure 2.6 Edge pixels determined by applying the chain code algorithm to the largest object in the breast mask of Figure 2.5(b).

Active contour models are invariably represented by a vector $\mathbf{v}(s) = (x(s), y(s))$, in terms of a normalised arc length parameter s. This parameter varies from 0 to 1 along the length of the snake contour. Kass *et al* defined a contour energy function

$$E_{snake} = \int_0^1 E_{int}(\mathbf{v}(s)) + E_{im}(\mathbf{v}(s)) + E_{ex}(\mathbf{v}(s))ds \tag{2.4}$$

where E_{int} represents the internal energy of the contour, E_{im} models interaction with the image, and E_{ex} models the external constraints. The internal energy of the snake can be further written as

$$E_{int} = (\alpha|\mathbf{v}_s(s)|^2 + \beta|\mathbf{v}_{ss}(s)|^2), \tag{2.5}$$

consisting of both first- and second-order derivatives. Adjusting the weights of α and β controls the relative importance of the first- and second-order terms and hence the extent to which the contour is allowed to stretch or bend respectively. For example, setting $\beta = 0$ allows the snake to become second-order discontinuous and develop a corner. Kass *et al* approximate these first- and second-order derivatives by finite differences. If $\mathbf{v}_i = (x_i, y_i)$ is a point on the contour, the approximations are calculated as

$$\left|\frac{d\mathbf{v}_i}{ds}\right|^2 \approx |\mathbf{v}_i - \mathbf{v}_{i-1}|^2, \tag{2.6}$$

and

$$\left|\frac{d^2\mathbf{v}_i}{ds^2}\right|^2 \approx |\mathbf{v}_{i-1} - 2\mathbf{v}_i + \mathbf{v}_{i+1}|^2. \tag{2.7}$$

To ensure the snake is useful for boundary refinement, energy functionals are required that attract the snake to salient features in the image. The total image energy E_{im} is expressed

as a weighted combination of these energy functionals,

$$E_{im} = w_{line}E_{line} + w_{edge}E_{edge}. \tag{2.8}$$

The simplest most useful energy functional is the image intensity itself,

$$E_{line} = f_i. \tag{2.9}$$

Depending on the sign assigned to w_{line}, the snake tries to align itself with the lightest or darkest nearby contour. Finding edges in the image can also be done with a very simple energy functional,

$$E_{edge} = -|\Delta f_i|^2, \tag{2.10}$$

where $|\Delta f_i|$ is the gradient magnitude of the grey level at position i. In this case the snake is attracted to contours with large image gradients. By adjusting the weights w_{line} and w_{edge}, a wide range of snake behaviour can be created.

The final term associated with E_{snake} is the external forces represented by E_{ex}. This is determined by external constraints imposed either by a user or a higher level process.

The snake energy E_{snake} is updated using an iterative process. During each iteration, a neighbourhood of each point i is examined and the point in the neighbourhood giving the smallest value for E_{snake} is chosen as the new location. Kass *et al* determines the minimum energy contour using a technique of variational calculus. Details of this minimisation procedure are presented in [79].

A number of difficulties arise when using the algorithm proposed by Kass *et al*. For example, there is no constraint for the inter-distance of the points on the contour. Therefore, when trying to find the optimal contour, points tend to bunch or overlap at lower energy fields. Kass *et al* also requires an estimate of high-order derivatives, which can lead to unpredictable behaviour of the contour. Amini *et al* recognised these problems and proposed an alternative time-delayed discrete dynamic programming algorithm for minimising the snake energy [2,3]. This allows addition of hard constraints, separate from the overall energy measure, providing more effective control over the snake behaviour. However, the method proposed by Amini *et al* is very slow, $O(nm^3)$, where n is the number of points on the contour and m is the size of the neighbourhood within which points are allowed to move at each iteration.

An alternative approach, known as the greedy snake algorithm [153], allows the inclusion of hard constraints as described by Amini *et al* but is much faster to implement, being $O(nm)$. The greedy algorithm is iterative, as are those of Kass and Amini and was chosen as the best method for the breast border refinement described herein.

The quantity minimised by the greedy algorithm is

$$E_{snake} = \sum_i [\alpha E_{spread}(\mathbf{v_i}) + \beta E_{curv}(\mathbf{v_i}) + \gamma E_{im}(\mathbf{v_i})], \tag{2.11}$$

where the first and second terms are first- and second-order continuity constraints, corresponding to E_{int} in Equation 2.4. The last term measures an image quantity, in this case grey level, and is the same as the middle term of Equation 2.4. A term for external constraints has not been included. The parameters α, β and γ are used to balance the relative influence of the three energy terms.

Using $|\mathbf{v}_i - \mathbf{v}_{i-1}|^2$ for E_{spread} causes the contour to shrink, as minimising this quantity minimises the distance between points. This also contributes to the problem of points bunching up at lower energy fields. A term which encourages even spacing of the points satisfies the original goal of first order continuity without the effect of shrinking. Thus, the greedy snake algorithm uses

$$E_{spread}(\mathbf{v}_i) = (\overline{d} - |\mathbf{v}_i - \mathbf{v}_{i-1}|)^2, \tag{2.12}$$

where $\overline{d}$ represents the average distance between all points in the contour. At the end of each iteration a new value of $\overline{d}$ is computed. As in the method proposed by Kass *et al* a reasonable and quick estimate for the curvature is given by

$$E_{curv}(\mathbf{v}_i) = |\mathbf{v}_{i-1} - 2\mathbf{v}_i + \mathbf{v}_{i+1}|^2 . \tag{2.13}$$

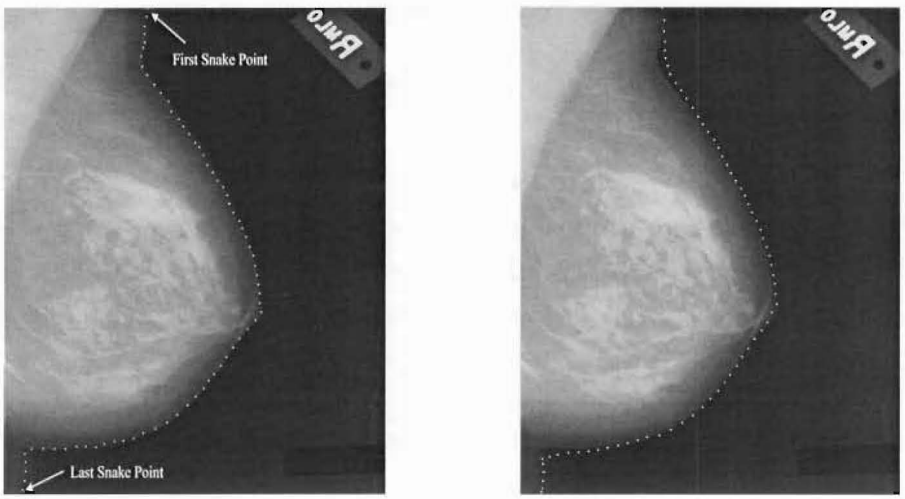
Both E_{spread} and E_{curv} model the internal snake energy and their combined influence ensures the contour does not shrink and remains smooth. The third term in Equation 2.11 is the image force, calculated as

$$E_{im}(\mathbf{v}_i) = f_i, \tag{2.14}$$

where f_i is the grey level value at pixel position i. This ensures the contour does not move towards places of high grey level, corresponding to points inside the breast border. Note that the image gradient has not been included in E_{im} because the gradient magnitude near the edge of the breast is not necessarily a maximum at the exact breast boundary.

Each energy term is normalised to within a range of 0 to 1, so that the relative magnitude of an individual term does not cause it to dominate [153]. Both the spread and curvature terms are normalised by dividing by the largest value in the current neighbourhood. The normalised value of E_{im} is obtained by dividing f_i by the maximum possible grey level value in the image (i.e. 256 for an 8 bit image).

For a current pixel location $\mathbf{v}_i$, and each of its eight nearest neighbours, the snake energy function E_{snake} is computed. The location having the smallest energy value is chosen as



(a) (b)

Figure 2.7 (a) Initial boundary snake points obtained from those shown in Figure 2.6. (b) Final boundary snake points after application of the greedy snake algorithm.

the new position $\mathbf{v}_i$. This process is applied to every point in the snake contour in turn to complete a single iteration. If any points along the contour move, the algorithm is repeated until convergence (to a local minimum) occurs. This is guaranteed since the cost measure is monotonically decreasing.

Application of the snake algorithm requires an initial contour to be placed near the object of interest. This is obtained from the chain-coded boundary described previously. Not all original points on the boundary are retained when implementing the snake algorithm as this would cause buckling to occur when the snake contour moves to a lower energy. It is however important to retain enough points on the initial boundary so that the breast edge detail is not lost. In this implementation every 17^{th} point in the chain code is retained for use in the greedy snake algorithm. This value was determined experimentally by testing a range of values (from 10 to 25) on FSD-A. Figure 2.7(a) shows the selected breast edge points from the boundary shown Figure 2.6. Two hard constraints are included in this implementation, ensuring the points marked as the first and last snake points are fixed, however all other selected points are allowed to move.

Optimal values for the three snake parameters were found to be

$$\alpha = 0.8, \qquad \beta = 1, \qquad \gamma = 1.2. \tag{2.15}$$

The parameters are not independent of one another and the optimum values were determined experimentally by applying the greedy algorithm to all images in FSD-A. Fig-

ure 2.7(b) shows the positions of the final snake points obtained using the above parameters. Nineteen iterations of the greedy snake algorithm were required before convergence occurred in this example. Application of this method ensures that many selected breast edge points, originally positioned inside the breast boundary, move to acceptable positions on the breast edge.

FSD-A was also used to test the algorithm's sensitivity to changes in the snake parameters, providing a measure of robustness. Two parameters were fixed at their optimal values while the third was allowed to vary, thus determining an acceptable range for the variable parameter, as follows:

$$0.5 \leq \alpha \leq 1 \qquad 0.8 \leq \beta \leq 1.2 \qquad 0.9 \leq \gamma \leq 1.5. \tag{2.16}$$

Factors considered when determining the parameter ranges included the bunching of points, and smoothness and location of the final breast boundary. Acceptable values for each snake parameter vary considerably (by a minimum of 20%) from the optimal values, indicating this method is robust to changes in parameter settings.

Spline interpolation

The last stage in the breast boundary detection algorithm involves application of a smooth interpolating function passing through the final snake points. This subsection firstly provides a description of the interpolation using a single dimension (1-D), which is then expanded to 2-dimensions (2-D) for application to the breast border representation.

To pose the problem in one dimension [33], suppose M node points (t_j, x_j) are given. For simplicity, assume

$$t_1 < t_2 < ... < t_M. \tag{2.17}$$

A function $q(t)$ is required, defined on $[t_1, t_M]$, that interpolates the data

$$q(t_j) = x_j \quad \text{for} \quad j = 1, ..., M. \tag{2.18}$$

The simplest method of interpolation is nearest neighbour interpolation, where the value at each position is set equal to the value of the nearest node point. Although this technique is simple to implement, it often produces a jagged output and has limited use in practical applications. Piecewise linear interpolation ensures points are connected by straight line segments. Once again however a smooth final output is not obtained. A further alternative is polynomial interpolation, creating an interpolating polynomial of degree $M-1$ [9]. This scheme provides a smooth interpolation function, however there is likely to be unacceptable oscillation between interpolating node points.

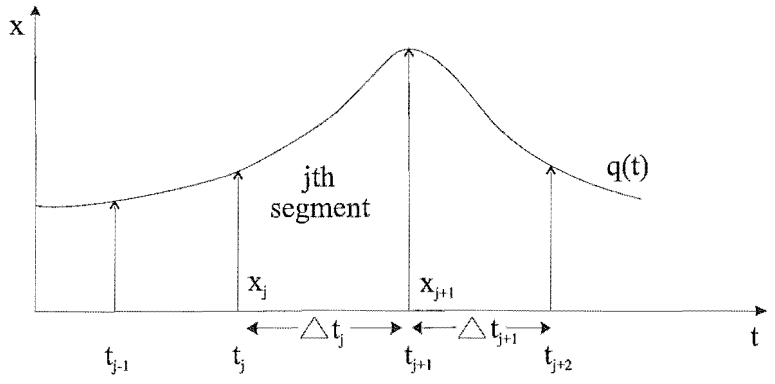


Figure 2.8 Example of cubic spline interpolation with four non-uniformly spaced knots.

Splines are piecewise polynomials consisting of pieces that are smoothly connected [33, 149]. The joining points of the polynomials, which have been referred to as node points above, are called *knots*. For a spline of degree K, each segment is a polynomial of degree K, which would suggest the need for $K+1$ coefficients to describe each segment. However, additional smoothness constraints are imposed on the continuity of the spline and its derivatives up to order $(K-1)$ at the knots, so that, effectively there is only one degree of freedom per segment. Splines of order 0 and 1 coincide with nearest neighbour and linear interpolants respectively. Increasing the order of the spline provides smoother interpolated functions, however oscillations are likely to occur. For this application a degree 3 polynomial is chosen, known as the cubic spline. This provides an acceptable trade off between smoothness and oscillation.

Figure 2.8 provides an example of a cubic spline with four non-uniformly spaced knots, where the value of Δt_j is calculated as

$$\Delta t_j = t_{j+1} - t_j, \qquad 1 \leq j \leq M-1. \tag{2.19}$$

Non-uniform spacing must be accounted for when applying spline interpolation to the final snake points on the breast boundary. The piecewise continuous cubic polynomial can be written as

$$q(t) = \sum_{k=0}^{3} a_{j,k}(t-t_j)^k, \qquad t_j \leq t \leq t_{j+1} \tag{2.20}$$

where $a_{j,k}$ are the polynomial coefficients. For a cubic spline the constraints within the jth

segment are

$$a_{j,0} = x_j, \qquad 1 \leq j \leq M-1 \tag{2.21}$$

$$x_{j+1} = \sum_{k=0}^{3} a_{j,k}\Delta t_j^k, \qquad 1 \leq j \leq M-1 \tag{2.22}$$

$$a_{j,1} = \sum_{k=0}^{3} k a_{j-1,k}\Delta t_{j-1}^{k-1}, \qquad 2 \leq j \leq M-1 \tag{2.23}$$

$$a_{j+1,1} = \sum_{k=0}^{3} k a_{j,k}\Delta {t_j}^{k-1}, \qquad 1 \leq j \leq M-2 \tag{2.24}$$

These constraints ensure that the resulting piecewise cubic function is continuous and has a continuous first derivative.

For this application nothing is known about the end point derivatives. Therefore the *not-a-knot* end condition is used which imposes a constraint on the slope at knot 1 and knot M so that the third derivative at knots 2 and $(M-1)$ is continuous [33]. Therefore,

$$a_{1,3} = a_{2,3} \tag{2.25}$$

$$a_{M-2,3} = a_{M-1,3} \tag{2.26}$$

which supplies the final two equations required to solve for the spline coefficients.

Before applying cubic spline interpolation to the breast outline, the final snake points are converted to their corresponding coordinates on the full resolution image. Conversion back to full resolution is performed prior to interpolation to avoid discontinuities in the final breast outline. A 1-D cubic spline function has been described above, however two dimensions must be taken into account for breast border interpolation. The total number of knots M corresponds to the number of final snake points including the first and last points as indicated in Figure 2.7(a). Each snake point has both an x and y coordinate value which are independent of one another. This ensures the 2-D interpolation can be split into two 1-D interpolations using a single dependent variable t. In this case t is the cumulative sum of the Euclidian distance (in 2-D) between consecutive snake points,

$$t_j = \sum_{i=2}^{j} ((x_i - x_{i-1})^2 + (y_i - y_{i-1})^2)^{1/2}. \tag{2.27}$$

Figure 2.9 shows plots of x and y versus t at the final snake points of Figure 2.7(b). Note that the knot points t are non-uniformly spaced. One dimensional splines are calculated for the two variables x and y using Equations 2.20 to 2.26, providing the 2-D coordinates of the final breast outline shown in Figure 2.10.

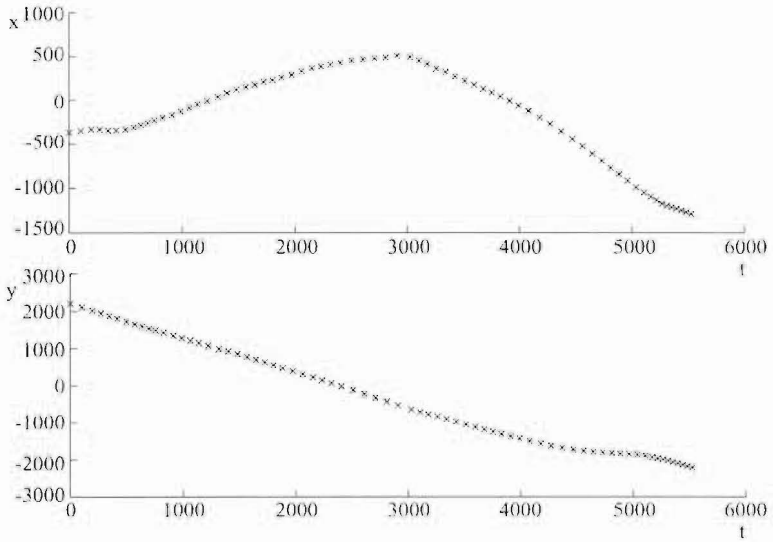


Figure 2.9 Two plots of x and y positions versus the cumulative Euclidian distances t between consecutive snake points.

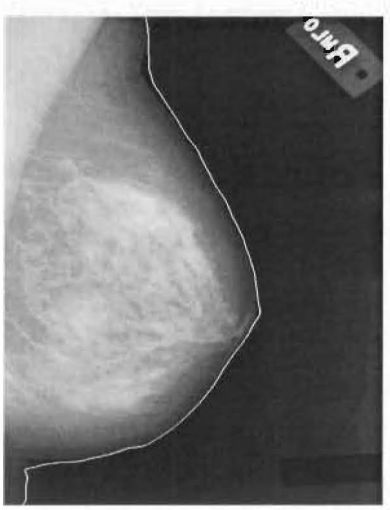
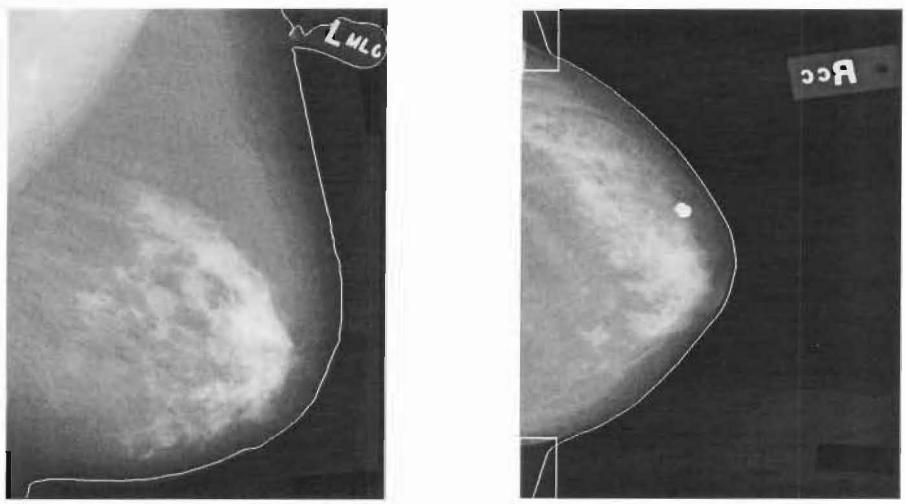


Figure 2.10 Final spline interpolated breast outline for the mammogram image in Figure 2.5(a).

Examples

The breast border detection scheme described above was trained and tested on FSD-A. Due to the lack of a ground truth for the breast tissue area the accuracy of the breast outline was determined by visual inspection. Results indicated that in 98% of the cases the segmentation algorithm produced a smooth breast outline, highly acceptable for CAD purposes. A single case produced an unsatisfactory breast border as shown in Figure 2.11(a). This was caused by the identification label being placed too close to the breast tissue area. At present the segmentation algorithm is unable to automatically determine when the breast border



(a) (b)

Figure 2.11 (a) Breast boundary obtained when the identification label is placed too close to the breast border in an MLO mammogram image. (b) Breast boundary indicating the higher intensity band of pixels present along the edge of a CC mammogram image.

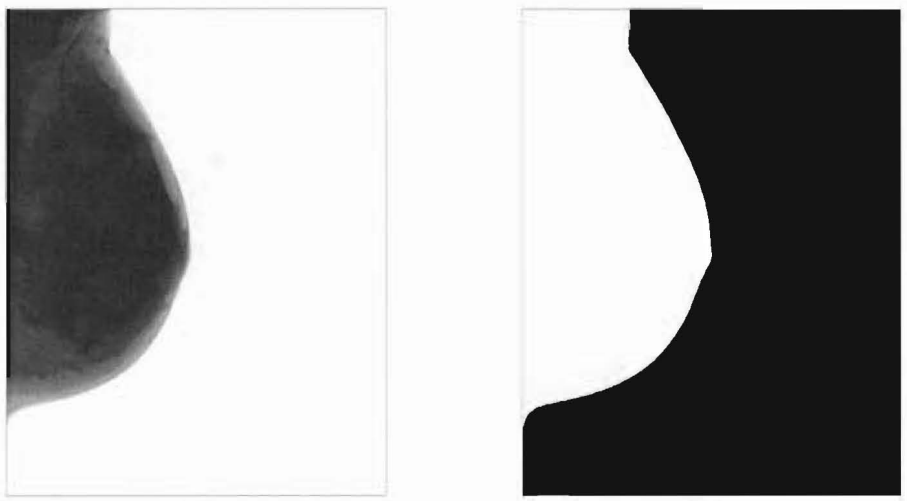
is incorrectly segmented. Other factors identified which are likely to cause unsatisfactory breast border detection include digitiser artifacts and poor mammographic technique, however these were not encountered in this data set.

In almost all images, a band of pixels with a higher signal intensity, up to $2cm$ in width, is found along the posterior edge of the direct-exposure area (see Figure 2.11(b)). It is suspected that these pixels are due to a tissue fold caused by imaging the breast as a two dimensional object. The phenomenon was also mentioned by Bick *et al* [14], however the author gave no reason for the increased signal intensity regions. Removal of these pixels requires additional processing, however the increase in computational load was deemed too great for the small decrease in processing time that would result in the subsequent computer programming stages. Therefore, these pixels remain classified as breast tissue pixels.

Direct digital mammograms

The segmentation scheme outlined in Figure 2.3, and described above, was developed for digitised film-screen mammogram images. A significantly simpler approach can be used for direct digital mammograms. As mentioned in Section 2.1, raw direct digital mammograms possess higher grey levels in the background compared to the breast tissue, an example of which is given in Figure 2.12(a).

When obtaining direct digital images the detector is saturated in the area outside the breast tissue, therefore all pixels in this area are set to a uniform pixel value. In an b-bit image,



(a) (b)

Figure 2.12 (a) Example of a 12-bit raw direct digital MLO mammogram image (b) Mask of the breast area after applying a global threshold value of 4095.

pixels associated with direct exposure (i.e. pixels that are not within the breast) are set to a grey level of $2^b - 1$. Additionally, there is no labelling required in a direct digital mammogram because this information is stored in a header file associated with each image. Therefore, segmentation of the breast area simply involves application of a global threshold. For the 12-bit mammogram image shown in Figure 2.12(a) a global threshold of 4095 is applied producing the binary breast mask in Figure 2.12(b). This global threshold was applied to the direct digital data set in Section 2.2 producing accurate breast masks in all cases. Obviously the processing time for breast area detection is greatly reduced when compared to digitised film-screen mammogram images.

2.3.4 Pectoral muscle

In the medio-lateral oblique (MLO) view of a mammogram the pectoral muscle appears in the top corner of the image. See for example the original mammogram in Figure 2.5(a) (MLO for the right breast). Knowledge of the pectoral muscle tissue and its boundary is a prerequisite for several tasks including determining mammogram adequacy and image registration. Although the pectoral muscle edge is not necessarily a straight line it is often modelled as one [42, 73, 88, 100]. The most common algorithm used for this application is the straight line Hough transform [37, 139]. This section outlines the principle of the straight line Hough transform and discusses its application for pectoral muscle boundary detection.

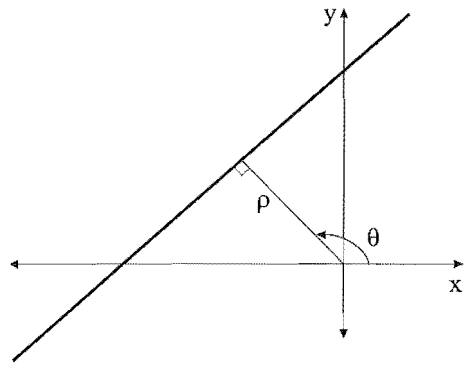


Figure 2.13 The normal parameters for a line.

Straight line Hough transform

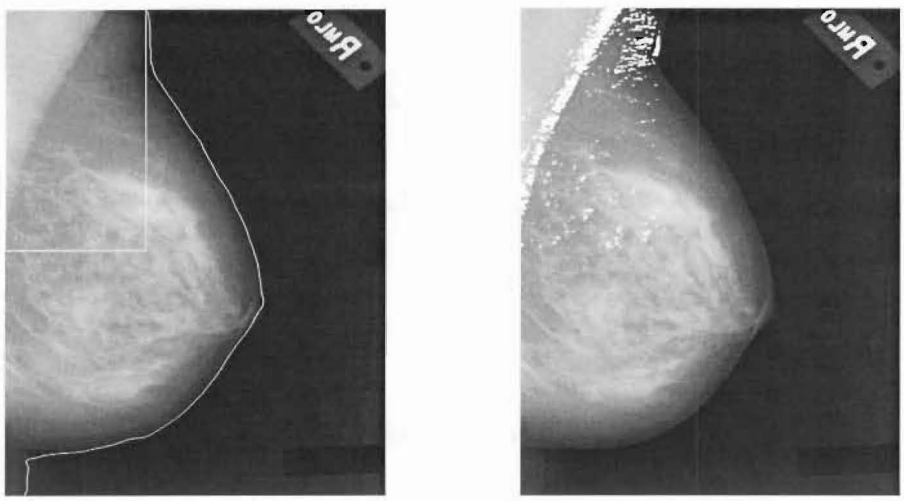
The Hough transform is a standard tool in image analysis that allows recognition of a global pattern in an image space by recognition of a local pattern (ideally a point) in a transformed parameter space [37]. The basic idea of the straight line Hough transform is to represent the set of points belonging to a single line in image space by a single point in parameter space. Figure 2.13 illustrates the line parameterisation chosen for pectoral detection. The parameterisation specifies a straight line by the angle θ of its normal and its algebraic distance ρ from the origin. The equation for a line corresponding to this geometry is

$$\rho = xcos\theta + ysin\theta. \tag{2.28}$$

A single point (x, y) in the image space transforms into a sinusoidal curve in the (ρ, θ) plane (Hough space), representing all possible lines that pass through (x, y). The curves corresponding to collinear image points have a common point of intersection in Hough space. This point in the Hough space uniquely defines the line passing through the collinear points. Thus the problem of detecting collinear points can be converted to the problem of locating intersecting curves.

When applying the straight line Hough transform to film-screen digitised MLO mammograms the following assumptions are made about the pectoral muscle [88]:

1. The muscle is located in the top left quadrant of the image, touching the top and left edges (assuming the breast is positioned with the chest side on the left of the image).



(a) (b)

Figure 2.14 (a) Film-screen MLO mammogram showing breast border and ROI for pectoral boundary detection. (b) Selected pixels (white pixels) for input to the Hough transform, after application of a gradient magnitude threshold.

2. The straight line estimating the pectoral boundary intersects the top and left edges of the image inside the breast region.
3. The intensity of the muscle is generally higher than that of the adjoining parenchyma.

Using this prior knowledge, a rectangular region of interest (ROI) is defined in which the pectoral boundary is searched for. The ROI extends from the top left corner, with height equal to half the image height and width corresponding to that at the top of the breast (see Figure 2.14(a)).

Inside the defined ROI the gradient magnitude $g(x, y)$ and direction $\phi(x, y)$ of the greyscale image $f(x, y)$ is calculated. Computation of the gradient is based on obtaining the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ at every pixel location. This is done by convolving two 7×7 Sobel operators with $f(x, y)$, the result being referred to as the *gradient image* [54, pp. 336-338]. To detect the pectoral outline the straight line Hough transform is performed on the gradient image. This image possesses high magnitude values at the pectoral boundary. The (ρ, θ) line parameterisation is used, as given in Equation 2.28, with the origin of the coordinate frame at the centre of the image.

The Hough space can be viewed as an accumulator array of (ρ, θ) values. Each pixel in the gradient image (within the defined ROI) increments those elements in the Hough accumulator array that represent the straight lines that pass through it. For pectoral boundary detection, constraints are applied to ensure only certain elements in the accumulator array are incremented. This not only avoids the selection of a straight line pectoral boundary that

is highly unlikely to be present in the mammogram image, but also decreases the execution time of the algorithm.

The first constraint is on angle. Each pixel in the defined ROI has a gradient orientation $\phi(x, y)$. This orientation is used to limit the range of the parameter θ such that

$$|\phi(x, y) - \theta| < \delta\theta. \tag{2.29}$$

A value of 0.03 radians was used for $\delta\theta$, experimentally determined using FSD-A.

The second constraint is based on gradient magnitude. A gradient magnitude threshold was set one standard deviation above the mean gradient magnitude of all pixels within the ROI (determined experimentally using FSD-A). Figure 2.14(b) illustrates the pixels selected (white pixels) for application to the Hough transform after gradient magnitude thresholding. The set of pixels selected include a high percentage of those on the pectoral boundary but typically only 10% of the total number of pixels in the ROI.

The increment value used in the accumulator array is determined by the gradient magnitude, giving pixels with a strong gradient a larger weight. Following [73], the weight function $w(g)$ is defined by

$$w(g) = \frac{\int_0^g h(g')dg'}{\int_0^\infty h(g')dg'} \tag{2.30}$$

with $h(g)$ the histogram of gradient magnitude values within the ROI. This definition makes the weight function invariant to monotonic transformations of the greyscale.

The assumptions mentioned previously on the pectoral muscle location and orientation, allow further constraints to be applied in searching for the peak in the Hough space. Thus the search is restricted to the region $\rho > 0$ and $0.7\pi < \theta < 0.98\pi$. Figure 2.15 shows an example of the accumulator array for the mammogram in Figure 2.14(a) after application of both the ρ and θ constraints (points outside the range are set to zero).

The final step of marking the pectoral muscle in the image is trivial. The position of the peak within the Hough space search region is inverse transformed into image space. All pixels on the marked pectoral boundary must be within the breast tissue area. If this is not the case, the next highest value in the accumulator array is searched for and the process is repeated until a straight line is found, representing the pectoral boundary, that is completely inside the breast area. Figure 2.16 shows the final straight line representation of the pectoral muscle boundary using the accumulator array in Figure 2.15. All pixels above the line and inside the marked breast tissue area are labelled as pectoral muscle.

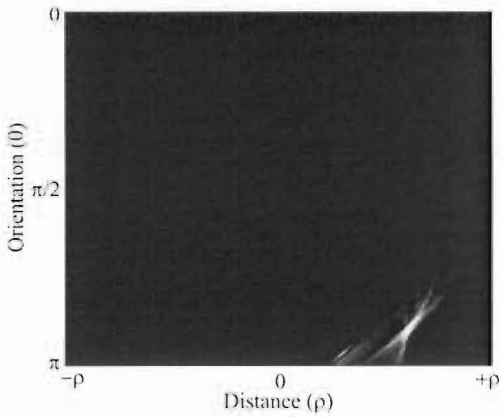


Figure 2.15 Accumulator array generated for the film-screen mammogram image in Figure 2.14(a).

Figure 2.16 Detected pectoral muscle using the accumulator array in Figure 2.15.

Examples

The straight line Hough transform was tested on FSD-A, containing 25 MLO views. A ground truth was not available for the pectoral muscle area therefore the accuracy of the pectoral muscle outline was determined by visual inspection. Acceptable results were obtained in 80% of the cases, three of which are shown in Figures 2.17(a-c). The most common cause for inaccurate pectoral muscle detection was the pectoral muscle boundary being curved (see Figures 2.17(d) and (e)). In these cases a straight line can only fit accurately to a portion of the muscle boundary. Figure 2.17(f) shows the result obtained when no clear boundary for the pectoral muscle can be seen. This only occurs as a result of poor mammographic technique. The major problem with inaccurate pectoral boundary detection arises when some breast tissue pixels are incorrectly classified as pectoral muscle. If all pixels classified as pectoral muscle are removed from subsequent computer programming

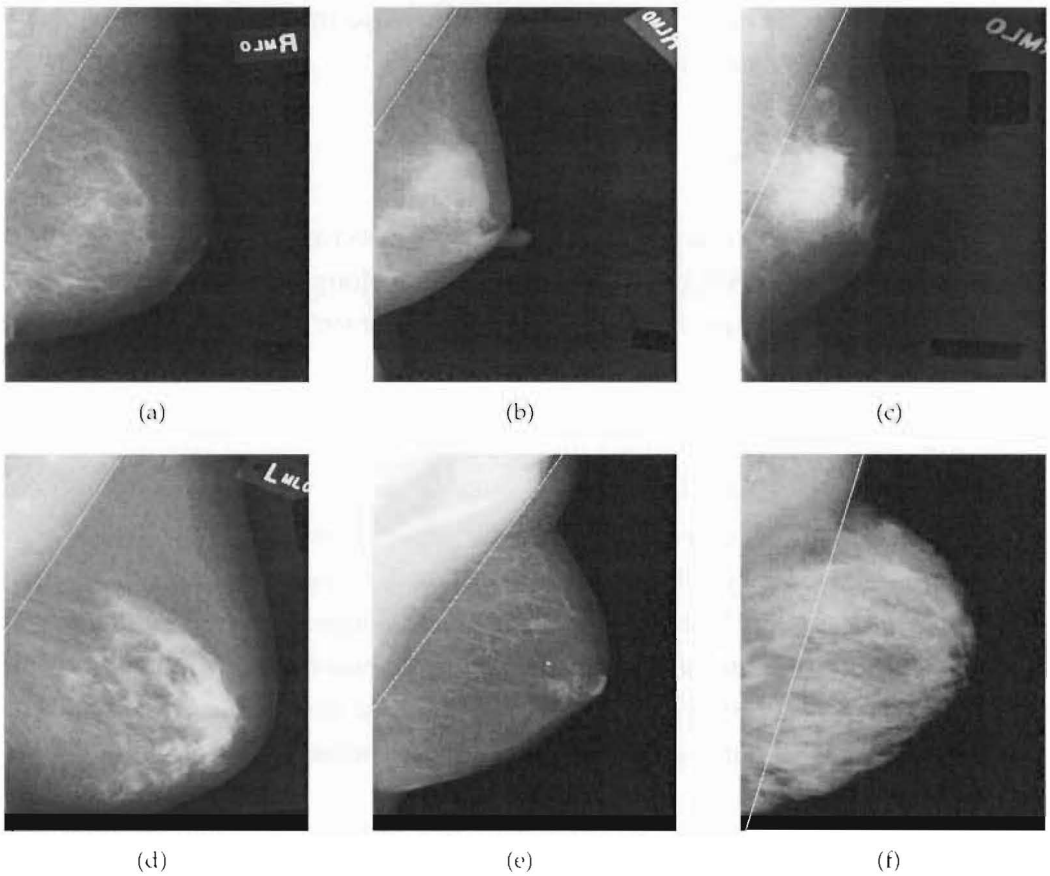


Figure 2.17 (a-c) Successful pectoral boundary detection using the straight line Hough transform using film-screen mammogram images (d-f) Poor pectoral boundary detection using the straight line Hough transform using film-screen mammogram images.

stages (restricting the search space for cancer detection), true positive detections could be missed due to the exclusion of a portion of the breast tissue area. At present this segmentation algorithm is unable to automatically determine when the pectoral muscle boundary is incorrectly segmented.

Alternative options for pectoral boundary detection were investigated in an attempt to improve the rate of acceptable detection. Firstly, the circular Hough transform [121, 139] was considered. This transform requires the determination of three parameters, and attempts to fit a circular arc for the pectoral muscle boundary. Results obtained using this algorithm were not sufficiently good to warrant the extra processing time required.

Also considered was the refinement of the original straight line pectoral boundary using the greedy snake algorithm [153]. This provided slight improvement when a good initial estimation of the pectoral boundary was found. However, when the initial estimate of the boundary was inaccurate (necessitating refinement of the outline), the snake failed to

locate the pectoral muscle edge. Once again the increase in execution time, for a small improvement in performance, was deemed unacceptable.

Direct digital mammograms

In a raw direct digital mammogram the intensity of the pectoral muscle is generally lower than that of the adjoining parenchyma. The straight line Hough transform described above can be adjusted for this change in image characteristic by simply changing the acceptable θ range to $1.7\pi < \theta < 1.98\pi$. All other stages of the Hough algorithm remain unchanged.

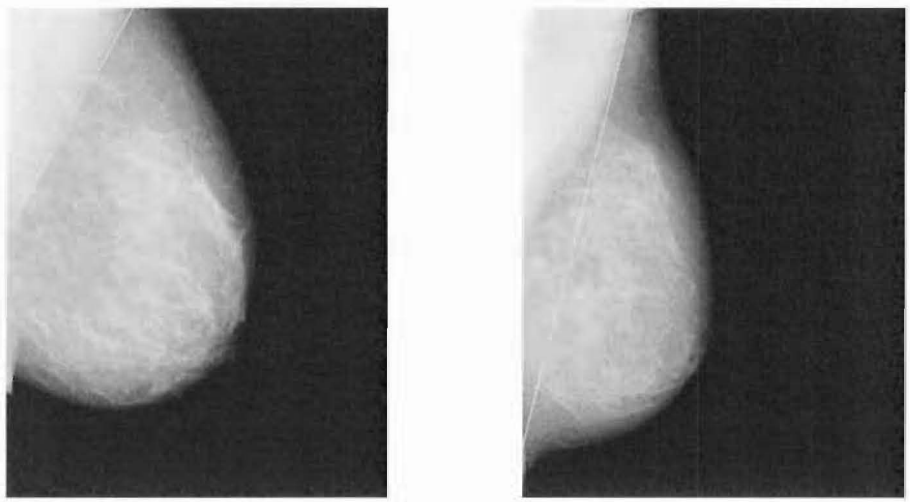
The transform was tested on the direct digital data set described in Section 2.2, comprising 62 MLO views. By visual inspection, acceptable results were obtained in 70% of the cases. Once again the most common cause for inaccurate pectoral muscle detection was the pectoral muscle boundary being curved. Additionally, two of the raw direct digital images provided a straight line boundary estimate inside the pectoral muscle due to a strong internal edge within the pectoral region. The pectoral boundary detection algorithm was trained using FSD-A and tested on the direct digital data set. Achieving a result of 70% accuracy for the direct digital mammograms is very encouraging considering the method is not necessarily optimal for this modality.

Figures 2.18(a) and (b) provide examples of both a good and poor pectoral boundary detection respectively for two direct digital mammogram images. Although the Hough transform was applied to the raw direct digital mammograms, for the purpose of display these images have been processed to ensure they are visibly similar to film-screen mammograms.

2.4 Summary

This chapter provided background information on the image characteristics of both film-screen and direct digital mammograms. Film-screen images possess a strongly nonlinear characteristic curve when compared to the direct digital modality. It is important when developing a CAD system that the image properties are understood, as they influence each stage of the algorithm design process.

Section 2.2 outlined the publicly available databases used to facilitate research in computerised mammography screening. Two film-screen data sets were selected from the DDSM, the first of which was a general data set consisting of both mass and microcalcification cancer types. The second was designed specifically for testing computer aided microcalcification detection programs, consisting of only microcalcification clusters. A single direct digital data set (also designed for microcalcification detection programs) was obtained from images produced on the GE Senographe system at the Nijmegen medical centre. All



(a) (b)

Figure 2.18 (a) Successful pectoral boundary detection using the straight line Hough transform on a processed direct digital MLO mammogram image (b) Poor pectoral boundary detection using the straight line Hough transform on a processed direct digital MLO mammogram image.

three data sets are used for the testing and training of the microcalcification CAD stages detailed in this thesis.

Section 2.3 described the pre-processing steps required for application of a CAD system to breast cancer detection. Accurate and robust segmentation methods for both the skin-air boundary and pectoral muscle boundary were developed. A multi-stage algorithm provides detection of the breast boundary in film-screen mammograms, while the image properties of the direct digital modality allow a simpler global thresholding technique to be applied in locating the breast tissue/air boundary. Segmentation of the breast tissue significantly reduces the processing time required for subsequent computer programming stages.

Automatic determination of the pectoral muscle boundary is not only useful in restricting the search space for cancer detection, but also in determining mammogram adequacy as well as in mammogram-pair registration and comparison. The straight line Hough transform was trained on digitised film-screen images producing accurate detection of the pectoral muscle boundary in 80% of the cases. When testing the algorithm on a direct digital data set accurate detection is still found in 70% of the cases.

The preprocessing stages detailed in this chapter provide a robust method of identifying the breast tissue area. All subsequent stages of image processing, described in the remainder of this thesis, are restricted to these selected pixels, reducing the processing time required without eliminating essential information for microcalcification detection.

Chapter 3

Noise equalisation in digitised film-screen mammogram images

3.1 Introduction

Digital mammograms can be obtained in one of two ways; by digitising conventional mammographic films or by using a direct digital image acquisition system. A major problem with both these techniques is the reduced spatial resolution when compared to conventional film-screen mammograms. Currently, digital systems for acquisition and display are limited to approximately 2000×2000 pixels, corresponding to a pixel size of roughly 100μm. To date the resolving power of conventional mammographic film-screen systems is at least two times greater than that of a digital system [75, 78].

Various studies have been carried out to determine the minimum required spatial resolution for accurate detection of breast cancer (more specifically microcalcifications) [28,29,75, 129]. Results indicate that a pixel size of 100μm is small enough to represent the individual shapes of microcalcifications with a performance comparable to film-screen mammograms. At this pixel size it is not the spatial resolution but the noise that limits microcalcification detection. The high spatial frequency noise present in mammogram images ensures that isolated spherical microcalcifications smaller than 130μm cannot be detected when using conventional film-screen mammography [75]. In order to accurately detect microcalcifications therefore, measurement of the high frequency image noise is of crucial importance.

This chapter focusses on the development of a noise model, used for equalisation of the high frequency noise present in digitised film-screen mammogram images. The noise equalisation process involves rescaling the mammogram image to help separate the signals caused by microcalcifications from those caused by high frequency noise. The image

characteristics of direct digital mammograms differ significantly from film-screen, therefore development of a noise model for the direct digital modality is left till Chapter 4. Section 3.2 outlines the noise components present in digitised film-screen mammogram images, identifying the need for a complex noise model. Section 3.3 gives a review of the noise compensation techniques previously developed in the literature. An accurate adaptive approach for modelling the image noise is described in Section 3.4 based on a method developed by Veldkamp *et al*, where the high frequency image noise is estimated as a function of the image intensity [150]. This adaptive approach is optimised in Section 3.5 by investigating a number of alternative implementations for estimating the image noise. Finally, Section 3.6 provides illustrative examples of how the optimised noise equalisation method works when applied to mammograms containing microcalcification clusters. It should be noted that conclusions about the accuracy of the noise model are difficult without utilising the model in a performance evaluation of a microcalcification CAD scheme. For this reason references are often made in this chapter to the final results presented in Chapter 8, where a performance evaluation is carried out.

3.2 Digitised film-screen image noise

Two main components must be considered when describing the noise in a digitised film-screen image: film noise and digitisation noise [78]. The factors that influence these noise components can be determined by looking at the physical aspects of the imaging and digitisation process.

The three primary sources of noise in a film-screen system are

1. the limited number of incident x-ray quanta,
2. film granularity, and
3. structure of the fluorescent screen.

All radiological images contain random fluctuation due to the statistics of x-ray quantum absorption. This noise can limit the reliability of detection of small or subtle structures such as microcalcifications [157]. In addition, noise due to the structure of the fluorescent screen and the granularity of the film emulsion used to record the image compound this problem. An ideal imaging system would be *x-ray quantum limited*, meaning that x-ray quantum noise is the dominant source of random fluctuation. When a system is x-ray quantum limited, further reduction in the noise requires an increase in the number of x-ray photons incident on the imaging system, that is, an increase in patient dose. Generally, existing mammographic film-screen systems are not quantum limited, because at high spatial frequencies noise in the imaging system is dominated by film granularity and screen structure, not by the number of x-ray quanta recorded [11, 119].

Other noise components are introduced when digitising a film-screen image. Digitisation is performed using a charge coupled device (CCD) sensor consisting of an array of light-sensitive elements producing varying amounts of charge in response to the amount of light they receive. At low light levels the signal caused by thermal excitation reduces the accuracy of the sensor, introducing what is known as *thermal noise*. The fundamental limit on noise performance in the sensor array however is *shot noise*, which is directly related to the number of photon detection events [78].

After sampling the image with a CCD sensor an analogue to digital conversion of the continuous sensor output is performed. This conversion reduces the information content of the signal, introducing quantisation as an independent additive noise term [53]. Research has shown that film noise is much larger than sensor or quantisation noise for the range of intensities most important in mammography [78]. Even though this is the case, the components that constitute film noise ensure that digitised film-screen mammograms possess very complex noise characteristics.

3.3 Existing methods

Most microcalcification detection methods described in the literature use some form of noise dependent adaptive threshold that is locally determined [24–26, 114, 117]. This is important for dealing with variation of the noise level across the image. One of the most well known methods of adaptive thresholding, developed by Chan *et al*, uses a global grey-level threshold as an initial processing stage followed by a locally adaptive threshold, which varies with the standard deviation of the surrounding pixel values [24–26]. Initially a difference-imaging technique is applied using linear spatial filters, to enhance the signal from microcalcifications while suppressing the background structure of the breast. A global grey-level threshold is used on the filtered image, extracting possible microcalcification pixels. To reduce the number of false signals caused by image noise a local grey-level threshold is applied to each potential microcalcification site. This threshold is based on the statistics (mean and standard deviation) of the pixel values in a local region of the mammogram image. The threshold level is set equal to a multiple of the standard deviation plus the mean pixel value. Although reasonable microcalcification cluster detection performance has been reported using adaptive thresholding (true positive rate of 82% with one false positive detection per image), the method suffers from a significant drawback. Adaptive thresholds that are determined locally in small image regions are not only determined by the image noise, but are influenced by image structures like lines and edges as well. Therefore, in regions with a lot of image structure, thresholds are not adjusted optimally to the noise level and detection performance may deteriorate.

Other researchers avoid using locally adaptive thresholds by taking the signal dependency of noise into account. Karssemeijer developed a preprocessing step for microcalcification

detection in which images are rescaled to equalise the noise [71,72,78], a method that was later improved upon by Veldkamp *et al* [150]. This preprocessing step transforms the input image in which noise depends strongly on the signal level into an image with a homogeneous noise level. Such an approach for microcalcification detection has been found to be beneficial in a number of studies and is singled out for further analysis in the remainder of this chapter. Netsch and Peitgen use a Laplacian scale-space representation of the mammogram, which is subsequently thresholded to determine potential microcalcification sites [115]. The method is significantly improved by considering the statistical variation of the complex noise characteristics of the mammogram (performing noise equalisation). Maitournam *et al* use splines to model the trend in each mammogram, which is subtracted from the original image, providing enhanced microcalcification visualisation. The noisy resultant image is thresholded after a variance equalisation is done [98]. Poissonnier and Brady compare two noise compensation techniques; the first is a transformation providing a linear relationship between greyscale and film density, while the second is noise equalisation. They concluded that noise equalisation is a superior approach as it is derived directly from each mammogram, thus it is adaptable to varying imaging conditions [124].

Additional approaches are based on adaptive noise suppression and other hybrid wavelet adaptive signal enhancement methods [125,126,141]. A disadvantage of these latter methods is that they suppress the noise (or enhance the signal) locally without taking the signal dependency of the noise into account.

3.4 Noise equalisation

Detection of local image features (such as microcalcifications) in digital x-ray images is often hampered by the fact that the noise models being assumed are too simple. Noise characteristics of digitised films are complex and depend strongly on the grey level itself. If this dependency is not taken into account adjustment of the sensitivity of a feature detection algorithm can be very hard, if not impossible, as the noise will vary with the mean grey level at the position of the feature detector. As mentioned in the previous section, one way to overcome this problem is to use local adaptive thresholding methods. In regions which are more or less homogeneous such an approach may work well. However, problems arise in the presence of boundaries between regions, where estimation of the image statistics is difficult.

An alternative and less time consuming approach is rescaling of the image pixel values, prior to feature extraction, to a scale on which the noise level is constant [72,78]. To equalise the noise level the meaning of the term *noise* should be defined carefully. Due to the high spatial frequency content and small size of microcalcifications, the detection accuracy of these features depends heavily on the high frequency noise components present in the

mammogram. Therefore it is reasonable to relate the noise to the standard error of pixel values in a high-pass filtered version of the original image.

A straightforward method for estimation of the high frequency noise as a function of the grey level would be to record a number of uniform samples at different exposures. Estimation of the noise would then require the processing of each sample separately. A disadvantage of using such a model is that dependency on the film type and film development characteristics is introduced. An alternative approach is taken here based on a method developed by Veldkamp *et al* [150]. This method enables estimation of the high frequency noise characteristics from an arbitrary image, provided that this image covers the whole range of grey levels of interest and that its structure is such that there are many more pixels in homogeneous regions than there are near feature boundaries. It should be noted that the filters used in the following description and the corresponding window sizes are scaled for images of 100μm per pixel.

Veldkamp *et al* related noise to the standard error of a local contrast feature representing the original image data [71,72,150]. Local contrast c_i at site i is defined by Veldkamp *et al* as

$$c_i = f_i - \frac{1}{N}\sum_{j\in\delta_i} f_j, \tag{3.1}$$

where δ_i represents a neighbourhood of N pixels centred on and including pixel i. The term *local contrast* is used here to avoid confusion when comparing this work with that described in [150]. Note also that the symbol 'c_i' is used throughout to represent local contrast for simplicity of notation and to maintain consistency with the earlier published work.

The dependence of the noise on the grey level f is removed by rescaling the contrast feature using the standard deviation of the local contrast $\sigma_c(f)$ as a measure of the noise. Figures 3.1(c) and (d) show examples of the local contrast images calculated from the digitised film-screen mammograms in Figures 3.1(a) and (b) respectively. Only pixels associated with the breast tissue area of each mammogram are included in the contrast calculation, all other pixels are set to zero. The contrast values c_i were obtained using a 9×9 window for $\delta_i(N = 81)$. This window size was chosen after determining the average microcalcification width (5.7 pixels) and average distance between neighbouring microcalcifications (21.7 pixels) for all clusters in FSD-B. Choosing a window size less than that of an individual microcalcification results in a decrease in contrast between the microcalcification and the background. On the other hand, choosing a window size too large covers multiple microcalcifications, which also decreases the contrast in the cluster region. For this reason a window size of 9×9 was chosen as it was considered the best window size to represent the relevant noise for microcalcification detection [150].

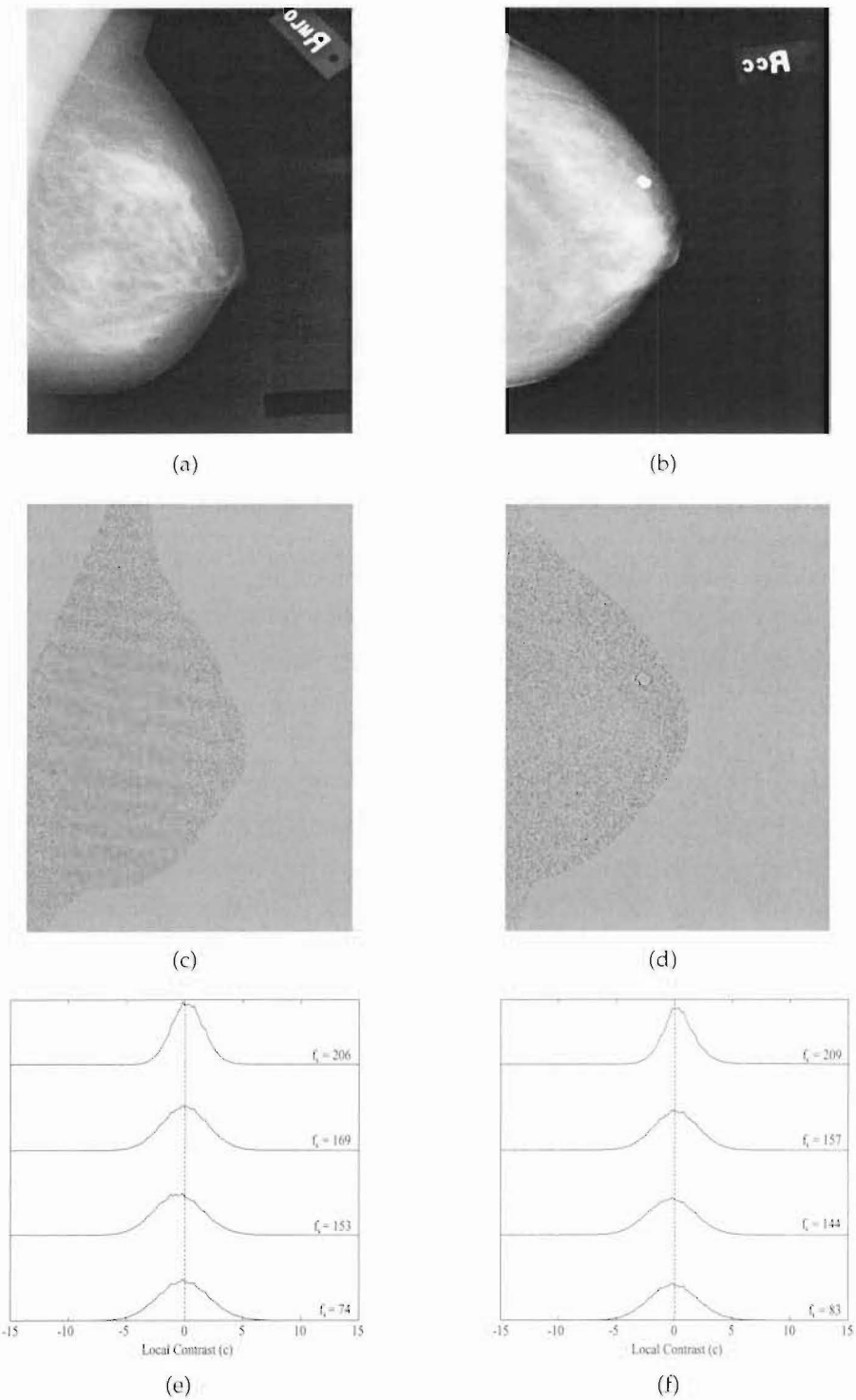


Figure 3.1 Digitised film-screen 8-bit mammograms showing two different views (a) MLO (b) CC. (c) Local contrast image for the mammogram in (a). (d) Local contrast image for the mammogram in (b). (e) Plots of $p(c|k)$ calculated from the mammogram in (a) for four different bins. (f) Plots of $p(c|k)$ calculated from the mammogram in (b) for four different bins. The vertical axis for each $p(c|k)$ represents the relative frequency of occurrence.

Equation 3.1 takes the form of a simple high pass spatial filter, therefore the standard deviation of the local contrast $\sigma_c(f)$ provides a measure of the high frequency image noise as a function of the grey level. To obtain $\sigma_c(f)$, the greyscale is divided into non-overlapping but adjacent bins numbered $k = 1, 2, \ldots, K$. After computing the local contrast, the probability density function $p(c|k)$ is estimated by normalising the histograms of c determined within each bin k. Figures 3.1(e) and (f) show examples of $p(c|k)$ (corresponding to four different bins) calculated from the digitised film-screen mammograms in Figures 3.1(a) and (b) respectively. Each bin k has a corresponding grey level f_k, calculated as the mean grey level of all pixels within bin k. The values of f_k are shown for each corresponding probability density function $p(c|k)$ in Figures 3.1(e) and (f).

Estimation of the standard deviation $\sigma_c(k)$ is performed by

$$\sigma_c(k) = \left(\int_{-\infty}^{\infty} (c - \mu_c(k))^2 \, p(c|k) \, dc \right)^{1/2} \tag{3.2}$$

where the mean local contrast for a particular bin $\mu_c(k)$ is

$$\mu_c(k) = \int_{-\infty}^{\infty} c \, p(c|k) \, dc. \tag{3.3}$$

Having obtained estimates for $\sigma_c(k)$ and $\mu_c(k)$, the continuous functions $\sigma_c(f)$ and $\mu_c(f)$ are estimated by interpolation. In this case a cubic-spline interpolation is applied, the details of which are provided in Section 2.3.3. Figures 3.2(a) and (c) show the continuous functions $\sigma_c(f)$ and $\mu_c(f)$ respectively, for the mammogram in Figure 3.1(a), with $k = 1, 2, \ldots, 16$. Similarly, Figures 3.2(b) and (d) show the continuous functions $\sigma_c(f)$ and $\mu_c(f)$, obtained using the mammogram in Figure 3.1(b). In these particular examples the number of pixels per bin is defined to be approximately equal. In common with Veldkamp *et al* [150] an extra point at the minimum grey level in the breast tissue region has been inserted to guide the interpolation along the lowest pixel values. The corresponding σ_c and μ_c values are set identical to the estimates $\sigma_c(f_1)$ and $\mu_c(f_1)$ respectively, representing the bin covering the lowest grey levels. This prevents the interpolation from producing erroneous results below f_1.

It is clear from these plots that $\sigma_c(f)$ varies strongly with the grey level. This variation is due to the complex characteristics of the noise components mentioned in Section 3.2. In both plots there is a rapid decrease in $\sigma_c(f)$ for the highest grey level values. This drop in noise level can be attributed to the decrease in the slope of the film's characteristic curve (see Figure 2.1(a)) [72].

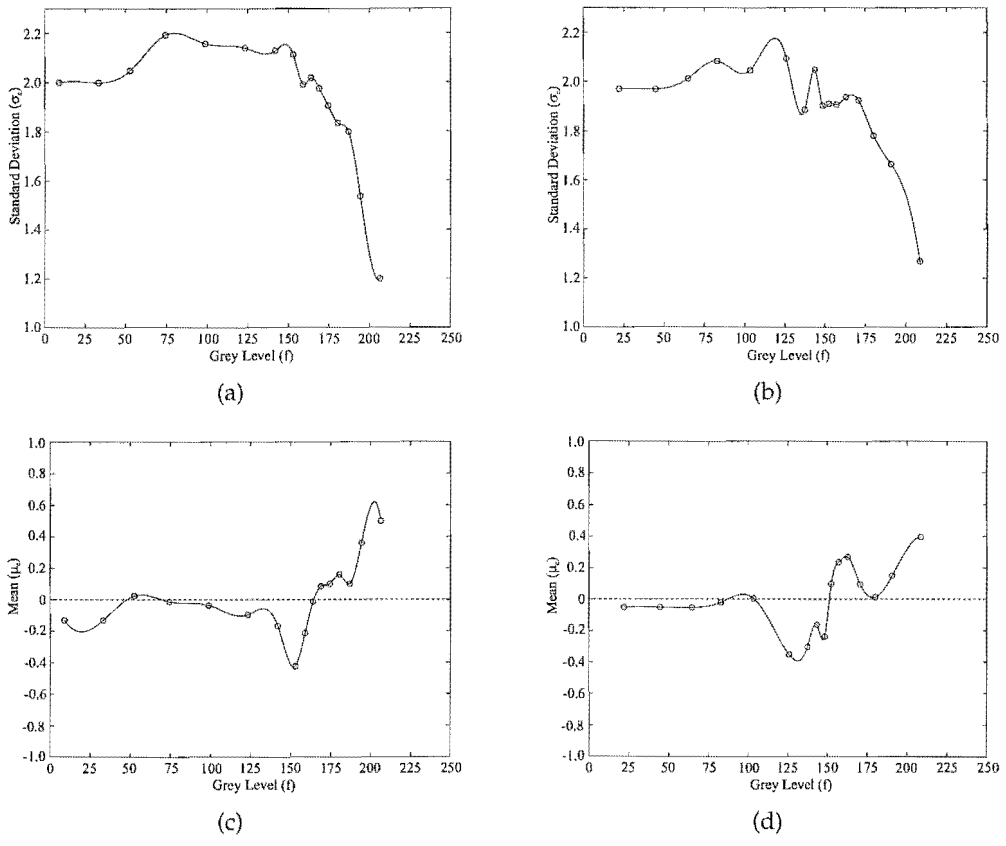


Figure 3.2 Contrast standard deviation $\sigma_c(f)$ and contrast mean $\mu_c(f)$ plots versus grey level. Left hand plots are for the mammogram in Figure 3.1(a), right hand for Figure 3.1(b).

The mean of each local contrast distribution $\mu_c(k)$ should be zero if the noise processes involved have a symmetric distribution, and if the curve relating the grey level to the x-ray exposure is approximately linear within each bin. However, it appears that small deviations of the mean around zero are quite usual. The effect of the bias can be partially explained using statistical probability. For example, the highest grey level pixels in a mammogram have a high probability of being surrounded by lower grey level pixels, ensuring they possess a positive mean local contrast value. In the lower grey level bins, however, the opposite effect occurs. Methods for compensating for this bias are described in subsection 3.5.3.

Statistical noise plots of both $\sigma_c(f)$ and $\mu_c(f)$ were obtained using the above technique, for all images in FSD-B. Although slight differences were observed between the plots over the set, the general trends mentioned above were observed in all cases.

The noise statistics $\sigma_c(f)$ and $\mu_c(f)$ are obtained for the purpose of equalising the image noise. By taking the signal dependency of both $\sigma_c(f)$ and $\mu_c(f)$ into consideration, the local

contrast c_i is equalised by

$$c_i' = \frac{c_i - \mu_c(f_i)}{\sigma_c(f_i)} \tag{3.4}$$

where c_i' represents the equalised local contrast at site i. This process removes the noise dependency on the grey level, and is known as *noise equalisation*.

3.5 Optimising noise equalisation

As explained in Section 3.1, detection of microcalcifications primarily concerns separating image signals caused by microcalcifications from those caused by high-frequency noise. An important step in the detection scheme, therefore, is the equalisation of local contrast. This section investigates a number of different implementations for local contrast equalisation aimed at improving the reliability of the noise statistic estimates. Firstly, a method of truncating the local contrast probability density functions prior to estimation of the noise statistics is described. Two variants for dividing the greyscale into bins are then investigated, the first using a fixed number of pixels per bin and the second using a variable number of pixels. As already discussed, histograms of local contrast are slightly biased, which causes non-zero mean values for local contrast. Methods that involve correction for the bias are investigated in an attempt to optimise noise equalisation.

3.5.1 Truncation

Local contrast equalisation of an arbitrary image relies on the fact that there are many more pixels in homogeneous regions than there are near feature boundaries. Contrast values associated with pixels near feature boundaries often possess very high magnitude, causing outliers to exist in the associated probability density function $p(c|k)$. Pixels associated with film artifacts contribute additional outliers, however these do not occur as frequently. For this reason, in the following explanation all outliers are assumed to be associated with feature boundary pixels.

Although, in a mammogram image, feature boundary pixels constitute a small proportion of the entire breast tissue area, their influence can significantly affect the estimates of both $\sigma_c(k)$ and $\mu_c(k)$. To ameliorate this problem, $p(c|k)$ is truncated to exclude extreme values assumed to be associated with feature boundary pixels. Estimates of the noise statistics $\sigma_c(k)$ and $\mu_c(k)$ are then determined based on the truncated distribution as follows.

All local contrast probability density functions $p(c|k)$ are assumed to have an underlying Gaussian distribution. This appears to be a valid assumption when looking at the plots in Figures 3.1(e) and (f). The purpose of truncating $p(c|k)$ is to estimate the mean and standard deviation of the underlying Gaussian and use these parameters as an estimate of

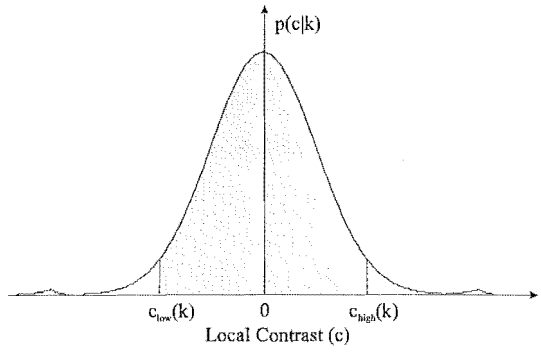


Figure 3.3 Local contrast probability density function $p(c|k)$ with corresponding truncation points $c_{\text{low}}(k)$ and $c_{\text{high}}(k)$ marked.

the noise level, thus ensuring the noise level is not influenced by the presence of outliers. Given the small number of pixels associated with feature boundaries in any given bin, the truncation points are chosen such that only a small proportion of $p(c|k)$ is excluded. Figure 3.3 shows an example of a local contrast probability density function $p(c|k)$ along with two truncation points $c_{\text{low}}(k)$ and $c_{\text{high}}(k)$. These truncation points are chosen to be symmetric around $c = 0$ and to satisfy

$$\int_{c_{\text{low}}(k)}^{c_{\text{high}}(k)} p(c|k)\, dc = L, \tag{3.5}$$

where L is a predefined fraction of the area under $p(c|k)$ (the shaded area in Figure 3.3). The relationship between $p(c|k)$ and the truncated probability density function $p_T(c|k)$ is defined as

$$p_T(c|k) = \begin{cases} \frac{p(c|k)}{L} & c_{\text{low}}(k) < c < c_{\text{high}}(k), \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

To estimate the mean and standard deviation of the underlying Gaussian based on $p_T(c|k)$, the sample mean $\bar{C}(k)$ and sample variance $s^2(k)$ of the truncated data are calculated, where $C(k)$ represents the local contrast of all pixels in bin k having values within the range $c_{\text{low}}(k) < c < c_{\text{high}}(k)$. These statistics are determined as

$$\bar{C}(k) = \frac{1}{M(k)} \sum_{i=1}^{M(k)} C_i(k) \tag{3.7}$$

and

$$s^2(k) = \frac{1}{M(k) - 1} \sum_{i=1}^{M(k)} (C_i(k) - \bar{C}(k))^2, \tag{3.8}$$

with $M(k)$ representing the number of pixels in bin k. A relationship is required between the known sample statistics of the truncated data and the population statistics of the underlying Gaussian. This relationship is found through the population mean $E[C(k)]$ and variance $\mathrm{Var}[C(k)]$ of the truncated distribution, defined as

$$E[C(k)] = \int_{-\infty}^{\infty} c\, p_T(c|k)\, dc \tag{3.9}$$

and

$$\mathrm{Var}[C(k)] = \int_{-\infty}^{\infty} (c - E[C(k)])^2\, p_T(c|k)\, dc. \tag{3.10}$$

Both $E[C(k)]$ and $\mathrm{Var}[C(k)]$ can be expressed in terms of the population mean, $\mu(k)$, and standard deviation, $\sigma(k)$, of the underlying Gaussian. Appendix B derives both expressions starting from Equations 3.9 and 3.10. For simplicity $\mu(k)$ and $\sigma(k)$ are replaced by μ and σ in all calculations. From Appendix B the population mean of the truncated distribution $E[C(k)]$ is given by [68]

$$E[C(k)] = \mu + \frac{Z((c_{\mathrm{low}}(k) - \mu)/\sigma) - Z((c_{\mathrm{high}}(k) - \mu)/\sigma)}{\Phi((c_{\mathrm{high}}(k) - \mu)/\sigma) - \Phi((c_{\mathrm{low}}(k) - \mu)/\sigma)}\sigma, \tag{3.11}$$

where Z is the unit normal density

$$Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{3.12}$$

and Φ is the standardised cumulative distribution function

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-x^2/2} dx. \tag{3.13}$$

Additionally the population variance of the truncated distribution $\mathrm{Var}[C(k)]$ is given by [68]

$$\mathrm{Var}[C(k)] = \left[1 + \frac{((c_{\mathrm{low}}(k) - \mu)/\sigma) Z((c_{\mathrm{low}}(k) - \mu)/\sigma) - ((c_{\mathrm{high}}(k) - \mu)/\sigma) Z((c_{\mathrm{high}}(k) - \mu)/\sigma)}{\Phi((c_{\mathrm{high}}(k) - \mu)/\sigma) - \Phi((c_{\mathrm{low}}(k) - \mu)/\sigma)} - \left(\frac{Z((c_{\mathrm{low}}(k) - \mu)/\sigma) - Z((c_{\mathrm{high}}(k) - \mu)/\sigma)}{\Phi((c_{\mathrm{high}}(k) - \mu)/\sigma) - \Phi((c_{\mathrm{low}}(k) - \mu)/\sigma)}\right)^2\right]\sigma^2. \tag{3.14}$$

With the values $c_{\mathrm{low}}(k)$ and $c_{\mathrm{high}}(k)$ known, the maximum likelihood estimation of $\mu(k)$ and $\sigma(k)$ is found by equating the first and second sample and population moments of the truncated distribution, resulting in $E[C(k)] = \bar{C}(k)$ and $\mathrm{Var}[C(k)] = s^2(k)$ [68]. In practice

Table 3.1 Sample mean and standard deviation for three different data sets associated with $p(c|k)$.

Data	Sample Mean	Sample Standard Deviation
Underlying Gaussian data	0.604	24.834
Complete Data (including outliers)	1.573	30.192
Truncated data	0.677	24.846

this is achieved by numerical minimisation of

$$(E[C_k] - \bar{C}(k))^2 + (\text{Var}[C_k] - s^2(k))^2. \tag{3.15}$$

A built-in MATLAB [101] function is used to perform the nonlinear minimisation procedure. Initial values for both $\mu(k)$ and $\sigma(k)$ are obtained from the sample mean and standard deviation respectively, of $p(c|k)$. On average the minimisation converges in less than 140 iterations, providing estimates for $\mu(k)$ and $\sigma(k)$. A reliable estimate of the noise level in each bin is achieved by equating the noise statistics $\sigma_c(k)$ and $\mu_c(k)$ to the estimates $\sigma(k)$ and $\mu(k)$ respectively.

To illustrate the effect of truncation on the estimated mean and standard deviation of $p(c|k)$, a Gaussian probability density function of $\mu = 0.5$ and $\sigma = 25$ was simulated. Outliers were added with high contrast magnitude and low probability, representing the presence of pixels near feature boundaries. For the purpose of this example, the distribution of these outliers was heavily biased towards positive local contrast values. Figure 3.4(a) provides an example of the simulated $p(c|k)$. By setting the value of L to 0.95 (corresponding to 95% of the distribution within the truncation points) the truncated density function $p_T(c|k)$ shown in Figure 3.4(b) was obtained.

Table 3.1 summarises the calculated sample mean and standard deviation for three different data sets associated with $p(c|k)$. As expected, the underlying Gaussian data set has both mean and standard deviation close to the simulated values of $\mu = 0.5$ and $\sigma = 25$. However, statistics based on the complete data set are significantly affected by the presence of outliers, causing both the sample mean and standard deviation to be overestimated. Figure 3.4(c) provides a visual comparison between $p(c|k)$ and a Gaussian estimation based on the statistics of the complete data set. Clearly, the standard deviation is too high for this density function, while the mean has a positive bias compared to the expected value of 0.5. Table 3.1 also shows that the statistics based on the truncated data set are in close agreement with that of the underlying Gaussian in $p(c|k)$. This is verified by looking at the plot of a Gaussian based on the truncated statistics overlayed on $p(c|k)$ in Figure 3.4(d).

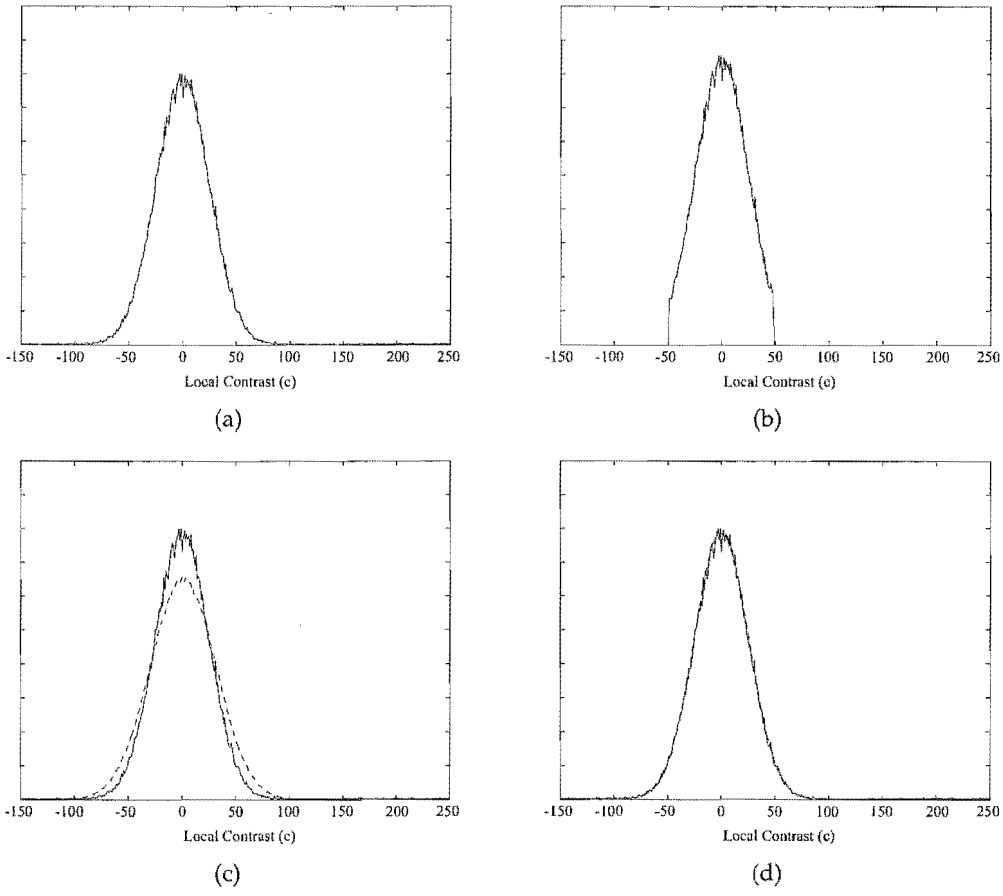


Figure 3.4 (a) Example of an original $p(c|k)$ (b) The density function $p_T(c|k)$ obtained from $p(c|k)$ using $L = 0.95$ (c) Gaussian estimation of $p(c|k)$ using the sample mean and standard deviation of the complete data. (d) Gaussian estimation using the sample mean and standard deviation of the truncated data.

The results presented here illustrate the importance of excluding outliers in $p(c|k)$ when obtaining the noise statistics $\sigma_c(k)$ and $\mu_c(k)$. The plots of $\sigma_c(f)$ and $\mu_c(f)$ shown in Figure 3.2 were calculated using the truncation method with $L = 0.95$. This value for L was determined after testing a range of values on all images from FSD-B ($0.80 < L < 0.98$). The influence of L on the noise statistic plots was found to be insignificant within this range. Setting $L = 0.95$ ensures a high percentage of $p(c|k)$ is included within the truncation points providing an good representation of the original probability density function without the inclusion of outliers. All results presented in the remainder of this chapter are obtained using this method of truncation with $L = 0.95$.

Quantitative results on the accuracy of the noise estimates when applying the truncation method to digitised film-screen mammograms are not available because of the complex noise model expected in this modality. More convincing proof on the improvement in noise estimation when using a truncated probability density function, $p_T(c|k)$, is however provided in the next chapter using direct digital step wedge phantom images.

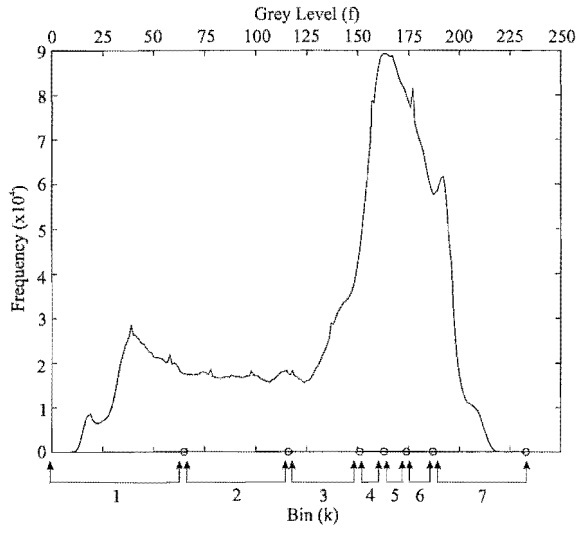


Figure 3.5 Grey level histogram of the breast tissue pixels of the mammogram in Figure 3.1(a). Also shown are the bin widths obtained using fixed size bins with $k = 1, 2, \ldots, 7$.

3.5.2 Determination of bins

Local contrast normalisation requires accurate estimation of the noise statistics covering the entire range of grey level values in the breast tissue region. Section 3.5.1 described a method of determining optimal values for $\sigma_c(k)$ and $\mu_c(k)$ based on the truncation of a given probability density function $p(c|k)$. These noise estimates are only valid, however, if $p(c|k)$ provides a good representation of the noise within bin k. Appropriate binning of the greyscale is therefore important, as it influences the reliability and accuracy of $p(c|k)$, and hence the reliability and accuracy of the statistics $\sigma_c(k)$ and $\mu_c(k)$.

This section investigates two approaches for binning the greyscale, the first of which uses non-overlapping bins of a fixed bin size, where the term *size* refers to the number of pixels per bin. The second method uses non-overlapping bins of variable size in an attempt to improve measurement of the noise statistics covering the highest grey level values [150].

Fixed number of pixels per bin

Dividing the greyscale into a predefined number K of fixed size bins means the widths of the bins are dependent on the histogram of grey level values in the breast tissue region. Figure 3.5 provides an example of a grey level histogram (of breast tissue pixels only) obtained from the mammogram in Figure 3.1(a). Shown at the bottom of this histogram are the widths of the fixed size bins when using $K = 7$. It is clear from this figure that the bin width is larger in sections of the greyscale corresponding to only a few pixels, whereas the bin width is smaller in sections consisting of a larger number of pixels.

Accurate measurement of the noise level requires determination of the appropriate number of bins K. When ascertaining this parameter, the following must be considered:

1. If the number of bins is chosen too small, significant changes in the noise level for a particular greyscale section could easily be missed due to each bin covering such a large range of grey level values.
2. If the number of bins is chosen too large, the estimated noise statistics will be inaccurate due to an insufficient representation of the local contrast values associated with each bin.

These factors indicate that the value of K must be sufficiently high to closely follow the change in noise level with respect to greyscale, without causing a decrease in accuracy due to a limited number of pixels per bin.

The contrast normalisation method described in Section 3.4 was applied to all images in FSD-B using the fixed size binning scheme. Figures 3.6 and 3.7 provide examples of the continuous functions $\sigma_c(f)$ and $\mu_c(f)$ obtained from estimates of $\sigma_c(k)$ and $\mu_c(k)$ respectively, for varying values of K. The plots on the left hand side correspond to the mammogram in Figure 3.1(a), while plots on the right hand side correspond to the mammogram in Figure 3.1(b). Once again a point at the minimum grey level in the breast tissue region, is inserted to guide the interpolation along the lowest pixel values.

Plots (a) and (b) in Figures 3.6 and 3.7 illustrate the continuous noise functions when using $K = 6$. These plots show the general trends in the noise level with respect to greyscale. Using such a small number of bins however, results in a noise model that is very dependent on the interpolation scheme and the small number of estimated $\sigma_c(k)$ and $\mu_c(k)$ values. Deviation from the true noise model is likely, due to the fact that each bin covers such a large range of grey levels.

At the other end of the scale, plots (g) and (h) in Figures 3.6 and 3.7 illustrate $\sigma_c(f)$ and $\mu_c(f)$ when using $K = 104$. These plots provide detailed changes in the noise level that are much less dependent on the interpolation scheme used. However, the small number of pixels associated with each bin cause inaccuracies in the noise estimates and thus fluctuations in the noise plots.

As mentioned in Section 3.5.1 quantitative results on the accuracy of the noise estimates are not available due to the complex noise model expected in this modality. For this reason, the ideal number of bins is determined based on visual comparison of the noise plots for varying values of K. After looking at the noise plots obtained from all images in FSD-B, for $3 < K < 150$, the optimal value of $K = 16$ was chosen. This value provides a good trade off between the dependency of the noise estimates on the interpolation scheme and

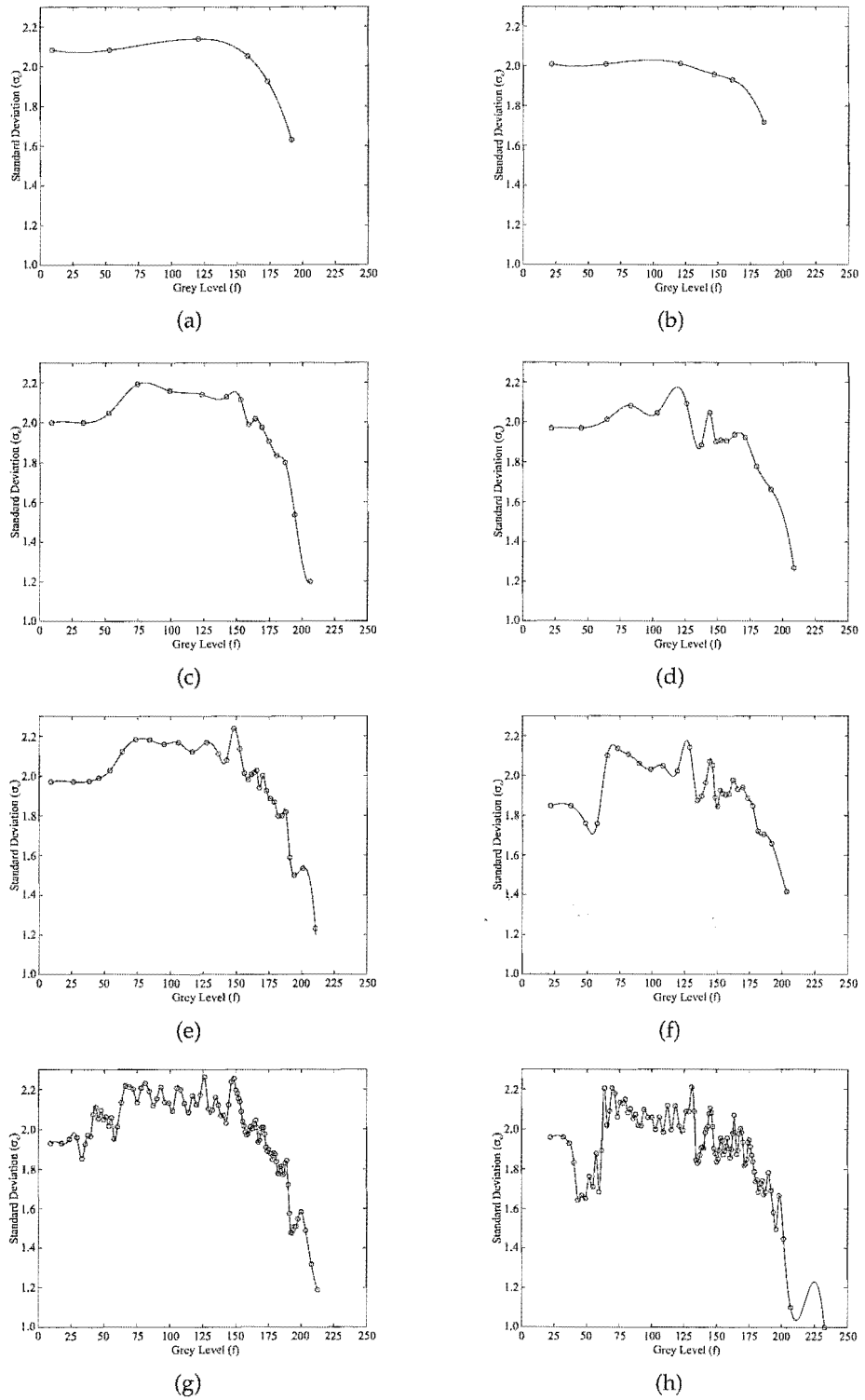


Figure 3.6 Plot of $\sigma_c(f)$ using fixed size binning with K equal to (a,b) 6 (c,d) 16 (e,f) 33 (g,h) 104. Left hand plots correspond to the mammogram in Figure 3.1(a). Right hand plots for the mammogram in Figure 3.1(b).

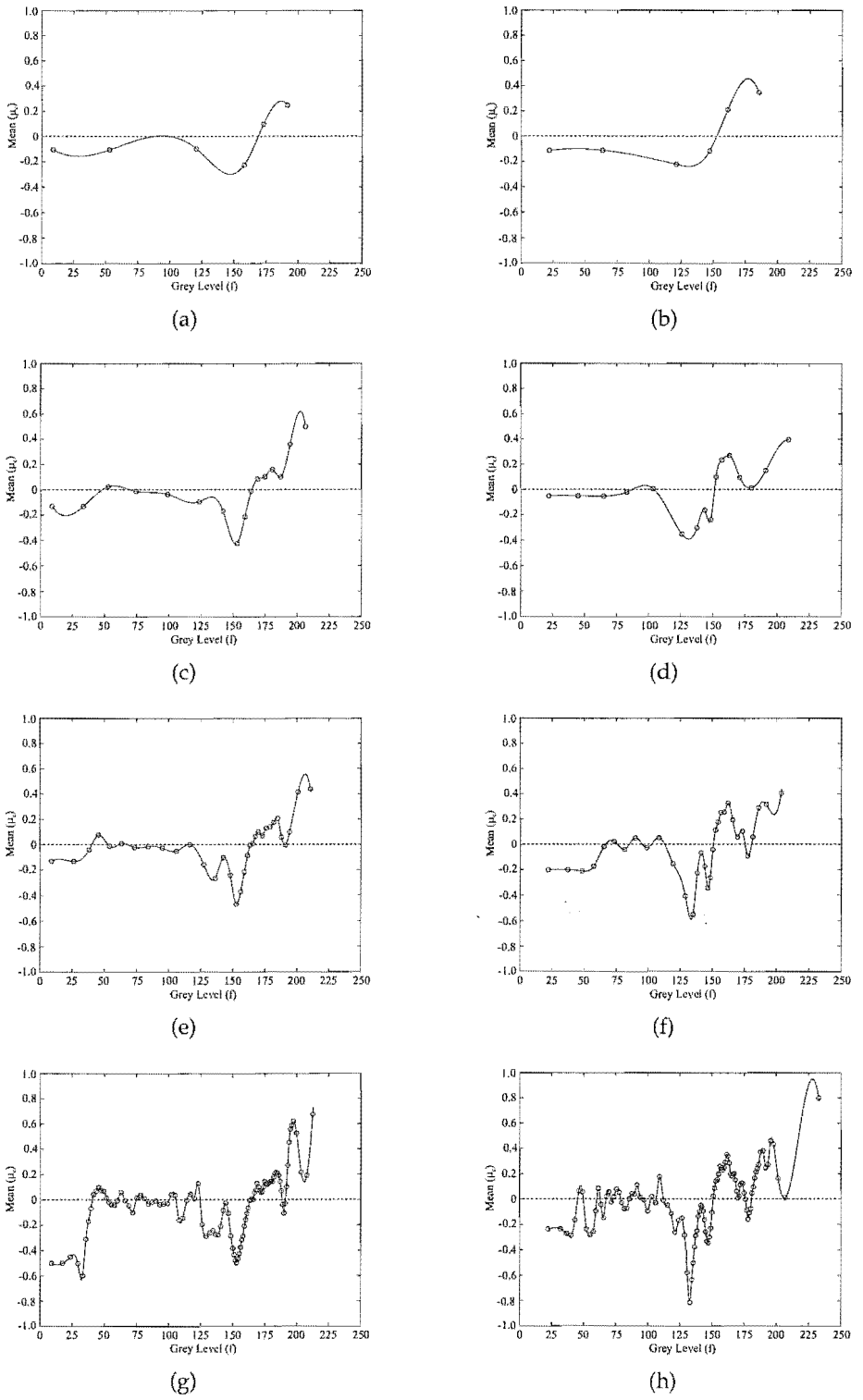


Figure 3.7 Plot of $\mu_c(f)$ using fixed size binning with K equal to (a,b) 6 (c,d) 16 (e,f) 33 (g,h) 104. Left hand plots correspond to the mammogram in Figure 3.1(a). Right hand plots for the mammogram in Figure 3.1(b).

the fluctuations in the noise level due to insufficient representation of local contrast values in each bin. Plots (c) and (d) in Figures 3.6 and 3.7 provide examples of the noise estimates when using $K = 16$.

Variable number of pixels per bin

It is difficult to obtain accurate noise estimates in the high intensity regions of a mammogram image. This is because the standard deviation of local contrast $\sigma_c(f)$ decreases rapidly with increasing intensity in the brightest regions. The rapid change in noise level is hard to measure because of the relatively small number of pixels that comprise these regions. However, accurate noise estimation in the brightest regions of a mammogram is important because microcalcifications often appear in these regions of the image. As a consequence of improper noise estimates in high intensity regions, a microcalcification detection scheme could miss true microcalcification clusters or generate relatively many false positive clusters.

More accurate estimates of $\sigma_c(k)$ are obtained in the brightest regions by using a variable number of pixels per bin $M(f_k)$. The value of $M(f_k)$ is chosen small for the highest grey levels because these regions consist of relatively few pixels. However, at lower grey levels, $\sigma_c(f)$ is more constant, and $M(f_k)$ is chosen larger. In this implementation, bins are defined to start from the maximum grey level in the breast tissue region. Thus avoiding a relatively small number of pixels remaining in the brightest region, which may be insufficient for estimating the noise accurately.

The first bin, covering the highest grey levels in the breast tissue area, consists of a small number of pixels M_{min}. The number of pixels $M(f_k)$ in each next bin k increases linearly until reaching a maximum bin size of M_{max} for bins covering the lowest grey levels (See Figure 3.8). The values of $M(f_k)$ are determined by the fraction of brightest pixels in the mammogram $Z(f_k)$, which is defined as

$$Z(f_k) = \int_{f_{k(max)}}^{\infty} p(f)df \tag{3.16}$$

where $f_{k(max)}$ is the highest grey level in bin k and $p(f)$ is the probability density function of grey level values f in the breast tissue region. The following relations describe the number of pixels per bin $M(f_k)$,

$$M(f_k) = \begin{cases} M_{max} & \text{if } Z(f_k) > Z_T, \\ M_{min} + \frac{Z(f_k)}{Z_T}(M_{max} - M_{min}) & \text{if } Z(f_k) \leq Z_T, \end{cases} \tag{3.17}$$

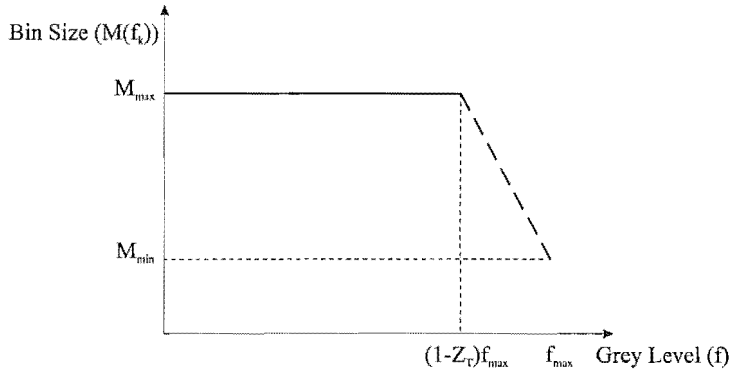


Figure 3.8 Size of each bin $M(f_k)$ with respect to grey level, f, when using the variable binning scheme. The right most portion is shown dashed as it is only approximately linear with respect to f.

where Z_T is a predefined fraction of the brightest pixels in the mammogram (discussed below). The parameters M_{max} and M_{min} are also predefined and are determined as a fraction of the total number of pixels in the breast tissue area. Figure 3.8 illustrates how the bin size $M(f_k)$ changes for varying grey levels, where the value of f_{max} is the maximum grey level in the breast tissue area.

There are three parameters to be determined when using the variable binning scheme, Z_T, M_{max} and M_{min}. In order to establish these parameter values, the motivation for variable binning is exploited. As mentioned previously, it is difficult to obtain accurate noise estimates in the high intensity regions of a mammogram due to the small number of pixels and the rapid decrease of $\sigma_c(f)$ in these regions. Therefore, it seems logical to design the variable binning scheme such that the bin size decreases at the point which the rapid fall of in $\sigma_c(f)$ occurs.

From the plot in Figure 3.8 the greyscale corresponding to the *knee* in the bin size is determined solely by the parameter Z_T. This greyscale value at which the bin size begins to decrease, f_T, differs for each digitised mammogram image. However, as long as good mammographic technique is employed, f_T only deviates by a small amount between images. By studying the $\sigma_c(f)$ plots obtained from FSD-B, when using the fixed binning scheme, the value of f_T was experimentally determined to equal 125 (see Figure 3.6). The value of Z_T is then calculated as

$$Z_T = \frac{f_{max} - f_T}{f_{max}}. \tag{3.18}$$

It was thought appropriate to set the number of bins K the same as that determined for the fixed size binning scheme ($K = 16$). This simplifies determination of the parameters M_{max}

Table 3.2 Values of M_{max} and M_{min} for the plots in Figures 3.9 and 3.10. Note that g and h are for fixed size bins ($K = 16$).

Plot	a	b	c	d	e	f	g	h
$M_{max} \times 10^3$	1703	1127	1216	805.2	912.2	603.9	380.1	251.6
$M_{min} \times 10^3$	0.060	0.040	6.081	8.052	24.33	20.13	380.1	251.6

and M_{min}. With a fixed number of bins, fixing the value of M_{max} results in a single value for M_{min} which satisfies the two constraints $f_T = 125$ and $K = 16$. Unfortunately, due to the fact that the bin boundaries are dependent on the grey level histogram of a particular image, derivation of a formula relating M_{min} and M_{max} is not possible. Therefore, each pair of M_{max} and M_{min} values are experimentally determined.

Once again the contrast normalisation method described in Section 3.4 was applied to all images in FSD-B, this time using the variable size binning scheme described above. Figures 3.9 and 3.10 provide examples of the continuous functions $\sigma_c(f)$ and $\mu_c(f)$ obtained from estimates of $\sigma_c(k)$ and $\mu_c(k)$ respectively, for varying M_{max} and M_{min} values.

The plots on the left hand side correspond to the mammogram in Figure 3.1(a), while plots on the right hand side correspond to the mammogram in Figure 3.1(b). For comparison with earlier results, Figure 3.9 and 3.10 repeat the results for the fixed size bins ($K = 16$) as the lowermost plots. Table 3.2 provides the exact values of the parameters used for the various plots in Figures 3.9 and 3.10.

From these plots it is clear to see that increasing M_{max} causes an increase in bin width and thus a less detailed noise plot in the lower grey level regions. Concurrently, decreasing M_{min} results in an increased number of bins in the higher grey level regions, thus a more detailed noise plot in the area where microcalcification clusters are likely to occur. However, for a fixed number of bins, there is a lower limit to the bin size beyond which the noise statistics become inaccurate. Plots (a) and (b) in Figures 3.9 and 3.10 illustrate that setting M_{max} too high results in a corresponding M_{min} value that is too small to sufficiently represent the noise level in the higher grey level bins. These unreliable noise measurements cause large fluctuations in the noise level which can clearly be seen in plots (a) and (b).

Plots (g) and (h) in Figures 3.9 and 3.10 show the noise plots obtained when M_{max} is set equal to M_{min}. This is equivalent to the fixed size binning scheme described in the previous section. Although fixed size binning provides reliable noise estimates covering the entire greyscale range, the accuracy of the estimates in the brightest regions can be improved by utilising the variable binning structure.

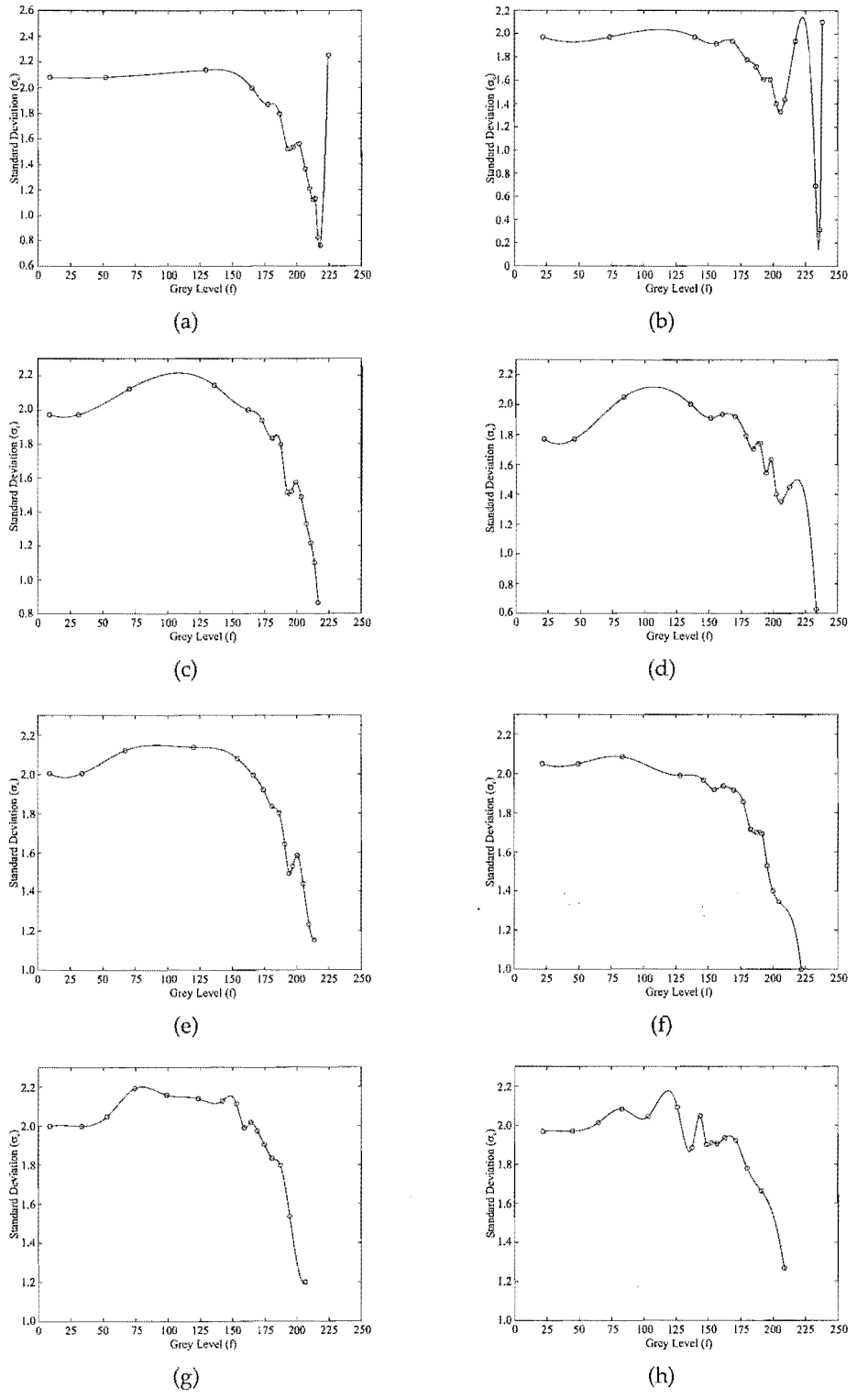


Figure 3.9 Plot of $\sigma_e(f)$ for the mammogram in Figure 3.1(a) (left) and Figure 3.1(b) (right) using variable size binning. See Table 3.2 for the parameter values used in each case.

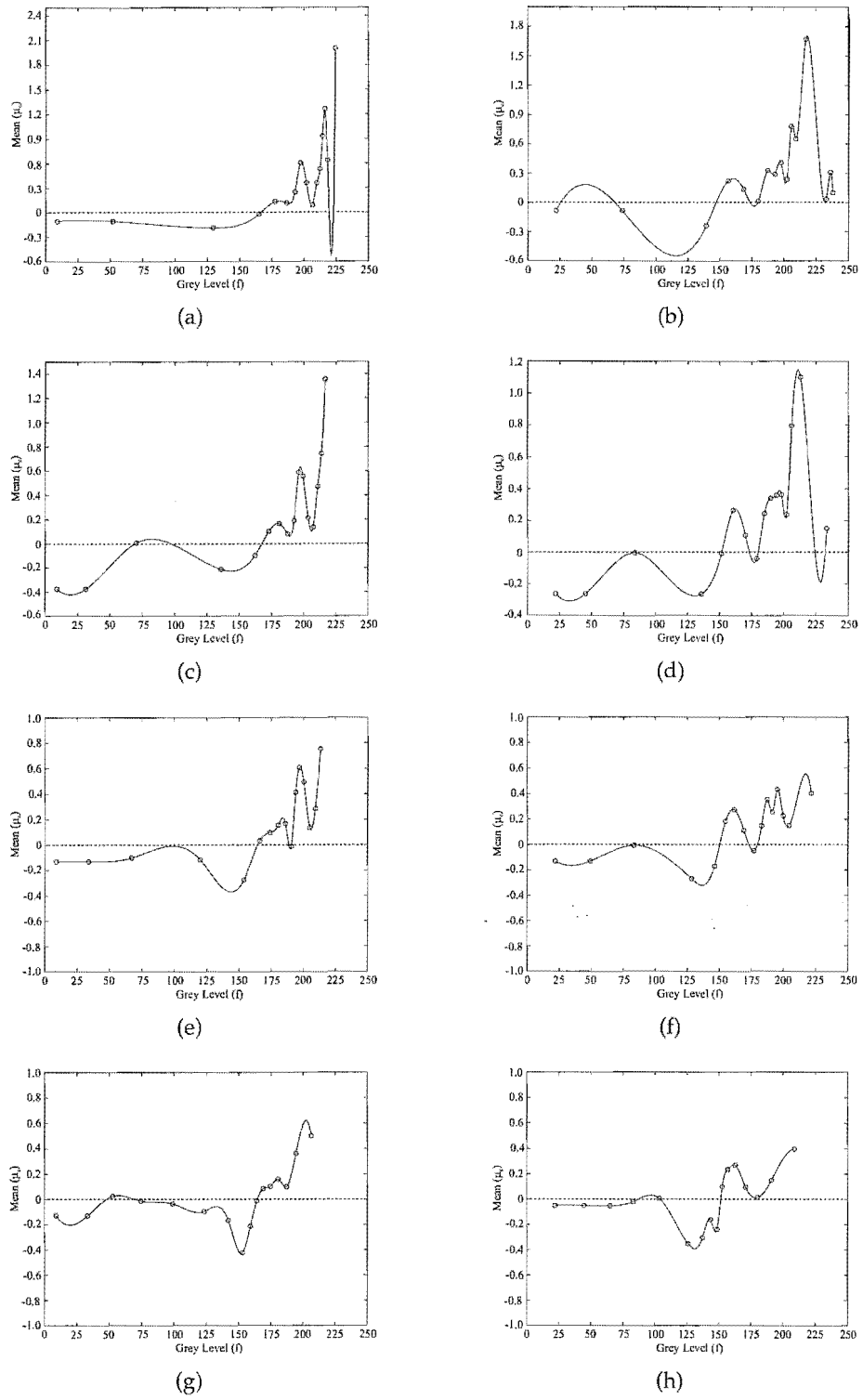


Figure 3.10 Plot of $\mu_c(f)$ for the mammogram in Figure 3.1(a) (left) and Figure 3.1(b) (right) using variable size binning. See Table 3.2 for the parameter values used in each case.

After visually comparing the noise plots for all images in FSD-B, using the variable size binning scheme, it was found that reliable noise measurements can be obtained for a large range of M_{max} and M_{min} values. Optimal settings for these parameters are difficult to obtain without knowledge of the exact noise level. Therefore, an arbitrary pair of M_{max} and M_{min} values have been chosen as being representative of the variable binning scheme ($M_{max} \approx 700 \times 10^3$ and $M_{min} \approx 22 \times 10^3$).

The logic behind variable binning suggests that it is a superior approach to the fixed size binning scheme. However, quantitative results on the accuracy of each method are not available due to the complex noise model that exists in this modality. It is hard to conclude therefore, which binning scheme is more appropriate without testing the microcalcification detection performance of a CAD scheme when using the two different binning techniques. This method of testing (using a CAD scheme) is carried out in Chapter 8, providing conclusive evidence that variable binning is superior to the fixed size binning method. Therefore, in the remainder of this thesis, when applying noise equalisation to digitised film-screen images, a variable binning scheme is used.

3.5.3 Bias in histograms of local contrast

It has been shown that the histograms of local contrast within each bin are not symmetric around zero, particularly if the number of pixels per bin is small. This is expected to affect noise estimation and thereby reduce the microcalcification detection performance of a potential CAD system. Three different methods of dealing with the bias in the local contrast histograms are investigated in this subsection. The first is to neglect the effect by setting $\mu_c(f)$ to zero, while two correction schemes attempt to reduce the bias. In the first correction scheme both $\mu_c(f)$ and $\sigma_c(f)$ are determined as a function of the grey level when performing local contrast equalisation. The second scheme reduces the bias in the local contrast histograms by changing the sampling scheme prior to local contrast equalisation.

Assuming $\mu_c(f)$ to be zero

When assuming the bias to be negligible, calculation of $\sigma_c(k)$ is obtained by setting $\mu_c(k)$ to zero in Equation 3.2. Local contrast equalisation is then carried out using

$$c_i' = \frac{c_i}{\sigma_c(f_i)}. \tag{3.19}$$

Assuming $\mu_c(f)$ to be grey level dependent

In this method, the mean local contrast value $\mu_c(f)$ is taken into account as a function of the grey level. Therefore, equalisation of the local contrast is performed according to Equation 3.4, where the standard deviation $\sigma_c(f)$ is determined with respect to $\mu_c(f)$.

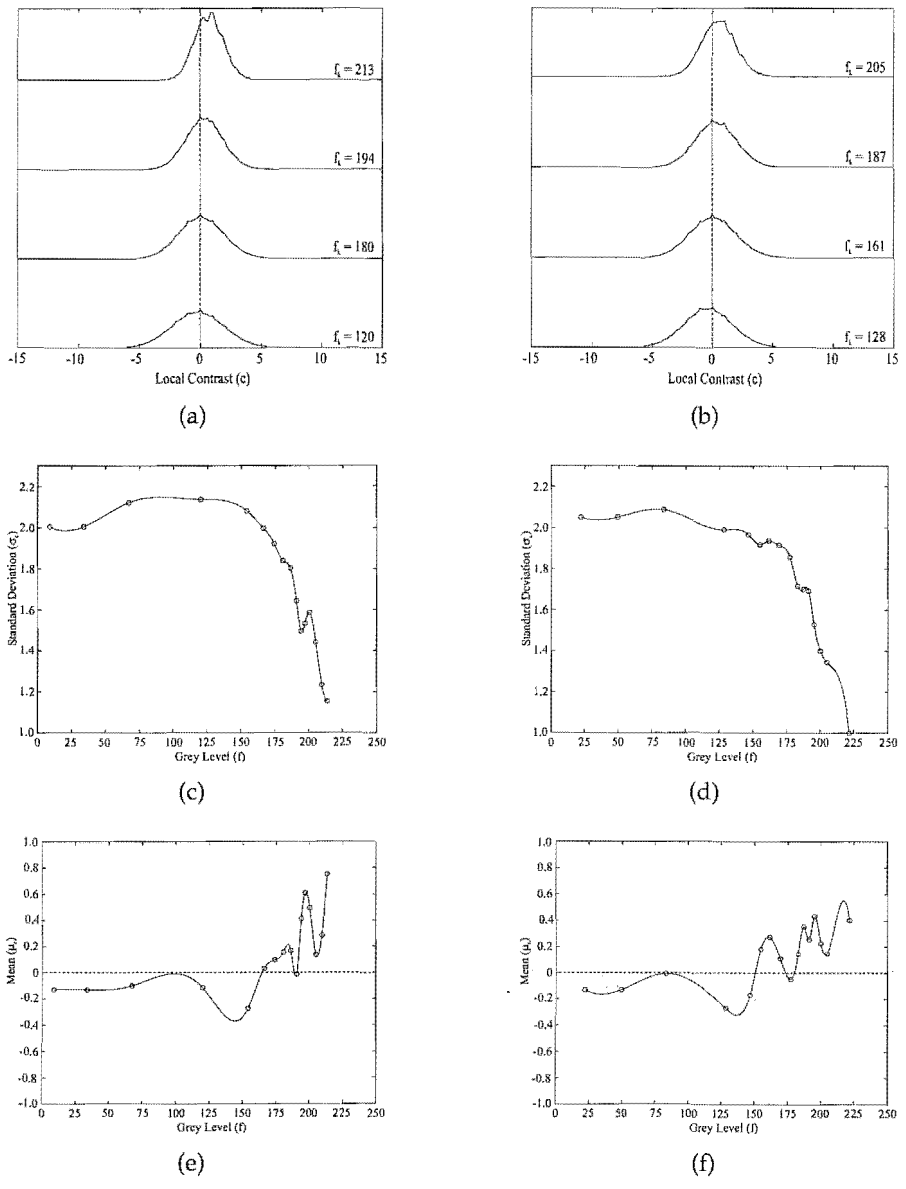


Figure 3.11 Histograms of local contrast and plots of both $\sigma_c(f)$ and $\mu_c(f)$ using the variable binning scheme based on f. Left hand plots are for the mammogram in Figure 3.1(a), right hand plots are for Figure 3.1(b).

Figures 3.11(a) and (b) show the bias that exists in the histograms of local contrast (for four different bins) from the mammograms in Figures 3.1(a) and (b) respectively. Also shown in these plots are the representative grey level values, f_k, for each bin. It is clear from these plots that in the lower grey level bins a slight negative bias exists, while in the higher bins a positive bias occurs. The reasons behind this effect have been explained in Section 3.4. Also shown in Figure 3.11 are plots of the standard deviation and mean of the local contrast as a function of the grey level. The plots of $\mu_c(f)$ reinforce the fact that variation of the mean around a local contrast value of zero is quite usual.

Reducing the bias in $\mu_c(\mathbf{f})$

In this method, an assumption is made that the relationship between a noisy pixel value f_i and pixel values without additional high-frequency noise components u_i can be written as [65]

$$f_i = u_i + \eta_i \tag{3.20}$$

where η_i is the high frequency noise contribution to u_i. In Section 3.4, calculation of c_i takes the form of a simple high pass spatial filter and is considered a measure for the high frequency noise contribution representing η_i. If for each u_i, the high frequency noise contribution c_i is known, then the noise level $\sigma_c(k)$ can be determined by dividing the greyscale into bins composed from the pixel values u_i and calculating $\sigma_c(k)$ from the local contrast distributions within each bin. By using u to bin the greyscale the noise processes within each bin are expected to be symmetric, and $\mu_c(k)$ is expected to be zero. Therefore, equalisation of the local contrast is performed using Equation 3.19.

Previously, c_i was calculated as the high frequency noise contribution to f_i instead of u_i. Therefore, a small error was made linking the local contrast values to the corresponding grey levels. According to Equation 3.20, selection of the appropriate bin for each pixel i should be based on the value of u_i. A simple estimate for u_i can be obtained by low-pass filtering f_i, using a mean filter denoted by

$$\hat{u}_i = \frac{1}{N} \sum_{j \in \delta_i} f_j \tag{3.21}$$

where δ_i represents a neighbourhood of N pixels centred on and including pixel i.

Binning with respect to $\hat{u}$ instead of f causes the reassignment of certain pixels to different grey level bins, resulting in a reduction in the bias. This reassignment can best be explained by looking at example pixels located at the extremes of a local contrast distribution $p(c|k)$. For a pixel at location i in an image, possessing a very high local contrast value, the relationship $\hat{u}_i < f_i$ exists. When binning with respect to $\hat{u}$ this pixel is likely to be assigned to a lower bin compared with binning based on f. On the other hand, for a pixel possessing a very low local contrast value, the relationship $\hat{u}_i > f_i$ exists. This pixel is likely to be placed in a higher bin when using $\hat{u}$ to bin the greyscale instead of f.

Figures 3.12(a) and (b) show the histograms of local contrast obtained when using $\hat{u}$ to bin the greyscale. The plot on the left corresponds to the mammogram in Figure 3.1(a) while the plot on the right corresponds to the mammogram in Figure 3.1(b). In this implementation $\hat{u}_i$ was obtained using a 9×9 window for $\delta_i (N = 81)$, which is equal to the window size used when calculating c_i. It is clear from these plots that the bias shown in Figures 3.11(a) and (b) has been reduced and the contrast distributions are now almost symmetric around $c = 0$.

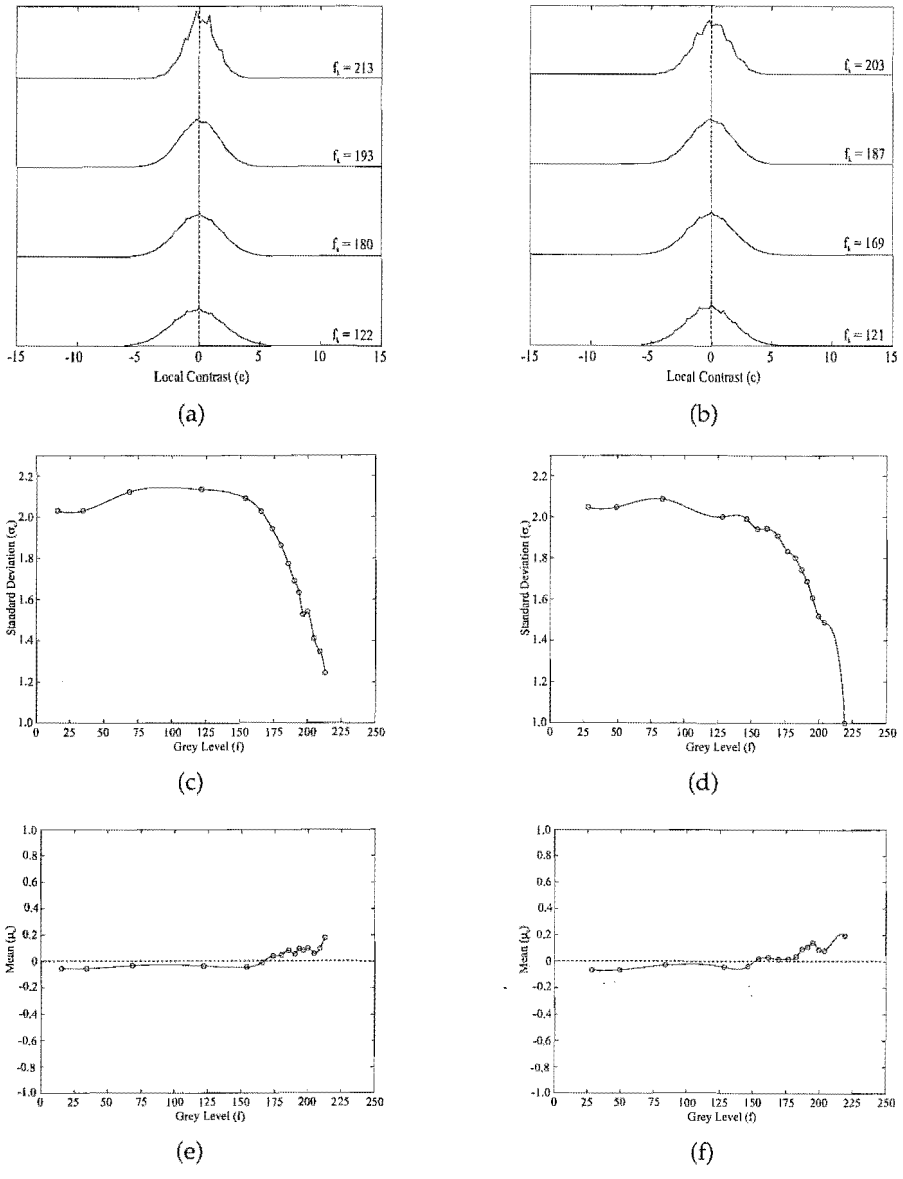


Figure 3.12 Histograms of local contrast and plots of both $\sigma_c(f)$ and $\mu_c(f)$ using a variable binning scheme based on $\hat{u}$. Left hand plots are for the mammogram in Figure 3.1(a), right hand plots are for Figure 3.1(b).

Figures 3.12(c) and (d) show the standard deviation of the local contrast as a function of the grey level. These plots are very similar to those shown in Figures 3.11(c) and (d), indicating that binning with respect to $\hat{u}$ does not considerably effect $\sigma_c(f)$. For the plots in Figures 3.12(e) and (f) however, a large decrease in variation of $\mu_c(f)$ is observed when compared to Figures 3.11(e) and (f). This provides further proof that binning with respect to a low-pass filtered version of the original mammogram does indeed reduce the bias of the local contrast distributions in each bin.

The effect of reducing the bias for noise equalisation can only be determined by testing the microcalcification detection performance of a CAD scheme using the three different techniques mentioned above. Chapter 8 carries out these comparisons, providing clear evidence that compensation for the bias improves CAD performance. Additionally, it is found that the two bias correction schemes, although providing an improvement in performance compared to assuming $\mu_c(f) = 0$, possess very similar CAD results. In the remainder of this thesis, when applying noise equalisation to digitised film-screen images, both $\sigma_c(f)$ and $\mu_c(f)$ are determined as a function of the grey level.

3.6 Equalisation examples

This section provides examples illustrating the results obtained after application of the noise equalisation method. Figures 3.13(a) and (b) show two sub-images, 480 × 480 pixels in size, from two digitised film-screen mammograms in FSD-B, both containing microcalcification clusters. Calculation of the local contrast using Equation 3.1 results in the images shown in Figures 3.13(c) and (d). Noise equalisation is then performed using the decided optimisation techniques mentioned in Section 3.5, resulting in an equalised local contrast image shown in Figures 3.13(e) and (f).

The local contrast images illustrate how equalisation of the image noise provides clearer visualisation of the microcalcifications present. In a potential CAD system, this ensures that separation of microcalcifications from background mammographic structure is much easier, thus improving the sensitivity of the CAD scheme. It should be noted that the equalised local contrast images not only highlight microcalcifications but also curvilinear structures (locally linear features). These features are often the cause of many false positive microcalcification detections due to the fact that they possesses similar properties (such as intensity and contrast) to microcalcifications. Detection of curvilinear structures within a mammogram image is described in detail in Chapter 5. Locating and removing these linear features helps improve the detection accuracy of a microcalcification CAD scheme.

3.7 Summary

Microcalcification detection performance is limited by the high spatial frequency noise present in film-screen mammogram images. In order to accurately detect microcalcifications, equalisation of the high frequency noise is of crucial importance. Section 3.2 provided a detailed description of the noise components associated with digitised film-screen mammograms, highlighting the need for a complex noise model.

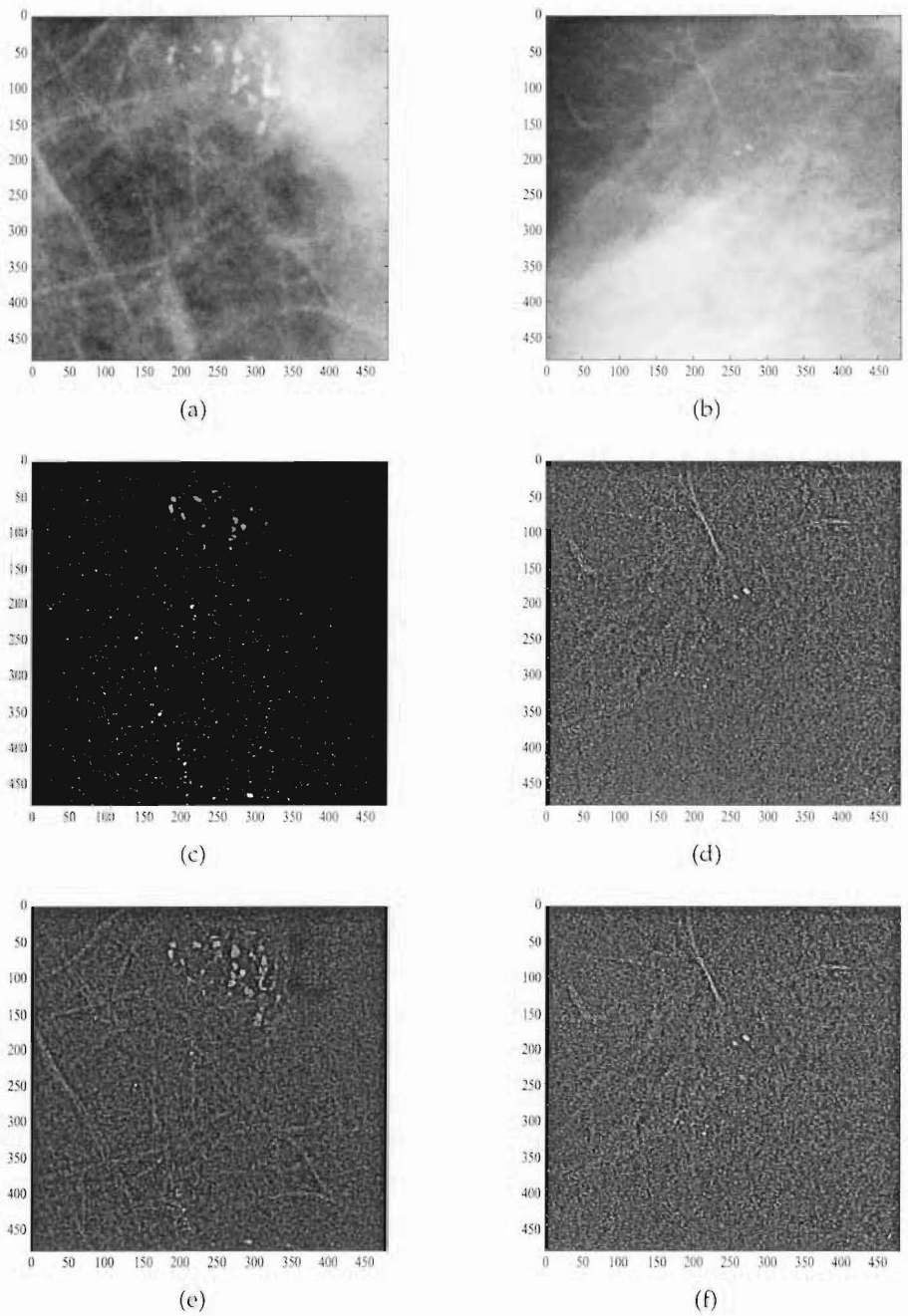


Figure 3.13 (a-b) Two sub-images from two digitised film-screen mammograms in FSD-B, both containing microcalcification clusters. (c-d) Local contrast image. (e-f) Equalised local contrast image.

As outlined in Section 3.3 many authors have attempted to compensate for noise when implementing microcalcification CAD systems. However, often the noise models being assumed are too simple, and the noise dependence on the greyscale is overlooked. Section 3.4 described an accurate adaptive approach for modelling the image noise based on a method developed by Veldkamp *et al.* In this method a high pass filtered version of the

original image, known as the local contrast, is used to represent the high frequency noise contribution associated with each image. The dependence of the noise on the grey level is removed by rescaling the contrast using the standard deviation of the local contrast as a measure of the noise. This approach is known as noise equalisation.

Section 3.5 investigated a number of different implementations for noise equalisation aimed at improving the reliability of the noise estimates. Firstly, a method of truncating the local contrast probability density function prior to estimation of the noise statistics was looked at. Truncation of $p(c|k)$ ensures that outliers associated with feature boundary pixels do not effect the noise estimation, providing a more reliable and accurate representation of the noise level.

Two variants for dividing the greyscale into bins were investigated, the first using a fixed number of pixels per bin and the second using a variable number of pixels. The variable binning scheme is designed to utilise the known general trend in the noise level with respect to greyscale. However, quantitative results on the accuracy of each method are not available due to the complex noise model that exists in this modality. Chapter 8 compares each binning scheme using the microcalcification detection performance of a CAD system. This testing provides conclusive evidence that variable binning is superior to the fixed size binning method.

Finally, methods of compensating for the bias in the local contrast distributions were investigated. Apart from neglecting the effect by setting $\mu_c(f)$ to zero, two correction schemes were looked at in an attempt to reduce the bias. In the first scheme both $\mu_c(f)$ and $\sigma_c(f)$ were determined as a function of the grey level. The second scheme reduced the bias in the local contrast histograms by changing the sampling scheme. Chapter 8 tests the methods of compensating for the bias on a CAD scheme. The testing shows that compensation for the bias improves CAD performance.

Chapter 4

Noise equalisation in direct digital mammogram images

4.1 Introduction

The motivation behind noise equalisation in direct digital mammograms is identical to that of the film-screen modality. When digital mammograms have pixels of 100μm or less, detection of microcalcifications is primarily limited by the presence of high spatial frequency noise [71]. Therefore, in order to accurately detect microcalcifications, measurement of the high frequency image noise is of crucial importance. In digitised film-screen mammograms a complex and highly variable relationship exists between intensity and noise variance. Direct digital images do not have this problem however, because a near linear relationship exists between pixel value and exposure (See Figure 4.1) [85, pp. 135-155]. This allows a simple noise model to be applied to direct digital mammograms, leading to a more robust estimation of the image noise.

This chapter focuses on the development of a noise model for direct digital mammograms based on the noise equalisation technique described in Chapter 3. Section 4.2 outlines the noise components present in direct digital mammogram images, identifying quantum noise to be the dominant noise source. Section 4.3 briefly mentions the noise compensation techniques previously developed in the literature. An accurate adaptive approach for modelling the image noise is described in Section 4.4 when quantum noise is dominant. In this case a simple square root noise model is used to estimate the high frequency image noise as a function of the image intensity. Experimental results are presented in Sections 4.5 and 4.6 when applying the noise model to direct digital step wedge phantom images and direct digital mammograms respectively. This is believed to be the first report discussing

the detection of microcalcifications in direct digital mammogram images, taking full advantage of the linearity of the system. As in Chapter 3 it is hard to make conclusions about the accuracy of the noise model for microcalcification detection without utilising the model in a performance evaluation of a microcalcification CAD scheme. For this reason references are made in this chapter to the final results presented in Chapter 8.

4.2 Direct digital image noise

All images generated by a quantised source are statistical in nature. Although the intensity at a point in the image can be predicted by the attenuation properties of the patient, the value fluctuates randomly about the mean predicted value [157]. The fluctuation of the x-ray intensity follows Poisson statistics, so that the variance, σ^2, about the mean number of x-ray quanta, N_o, falling on a detector element of a given area, is equal to N_o. In order to produce a signal the x-ray quanta must interact with the detector material. Interactions can either be photon scattering (the most common form of which is Compton scattering) or photoelectric absorption. Of the total x-ray signal about 40% is scattered radiation, however it is the photoelectric absorption that contains useful information for diagnosis [61, pp. 31-44]. This interaction can be represented as a binomial process with probability of success, η. It has been shown that if fluctuation in x-ray intensity is Poisson distributed then the distribution of interacting quanta due to photoelectric absorption is also Poisson distributed with standard deviation [12]

$$\sigma = (N_o \eta)^{1/2}. \tag{4.1}$$

Independent sources of noise may contribute at different stages of the imaging system. Their effect on the noise variance is additive and the fluctuation is subject to the gain of subsequent stages of the imaging system. An ideal imaging system would be *x-ray quantum limited*, meaning that the level of quantum noise determines the system performance. When a system is quantum limited reduction in the relative noise can only be achieved by an increase in the number of x-rays incident on the imaging system, that is an increase in patient dose. Achieving the best quality digital mammogram means balancing the need for keeping patient dose as low as possible with the desire for the least noise possible.

In a well designed direct digital system the number of quanta incident on the detector and η are the main factors influencing the signal-to-noise ratio [157]. Hence, following the Poisson model in Equation 4.1, a square root relationship exists between the noise and the number of x-ray quanta.

A direct digital system has a linear relationship between grey level and exposure (and, therefore a linear response to x-ray photons) as is shown in Figure 4.1. These plots were

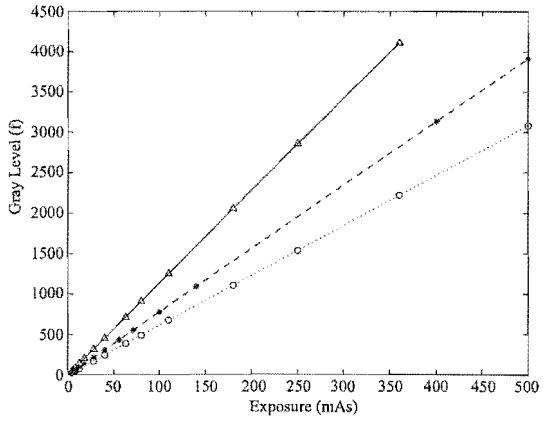


Figure 4.1 Experimental plots of pixel value (12-bit number) versus exposure in mAs for a 5.0 cm polymethyl methacrylate (PMMA) block at 28kVp and three target/filter combinations: △: rhodium/rhodium, ○: molybdenum/molybdenum, and ∗: molybdenum/rhodium. The mean grey level was calculated for a 4 cm^2 region-of-interest in the centre of the image.

obtained using the General Electric (GE) Senographe system at the Nijmegen medical centre. The acquisition parameters mentioned in this figure are described in detail in subsection 4.5.1.

The assumed quantum noise dominance of the direct digital system and the linear characteristic curve that this modality exhibits ensures that a simple square root noise model provides accurate estimation of the noise with respect to grey level.

4.3 Existing methods

Most microcalcification detection schemes described in the literature are developed for digitised film-screen mammograms. The noise compensation techniques used in these methods are therefore optimised for the film-screen modality (See Section 3.3). Although many of the noise compensation methods can easily be applied to direct digital images, the vast difference in characteristics between film-screen and direct digital images makes it more logical to develop a method specifically for direct digital mammograms based on the noise components present. Such a method is presented here and to the author's knowledge has not been previously addressed in the literature.

4.4 Noise equalisation

A noise equalisation scheme designed specifically for digitised film-screen mammogram images has been described in Chapter 3. The same principle is employed here, except that the expected square root relationship between the noise level and the local contrast is used, leading to a more robust estimation of the image noise.

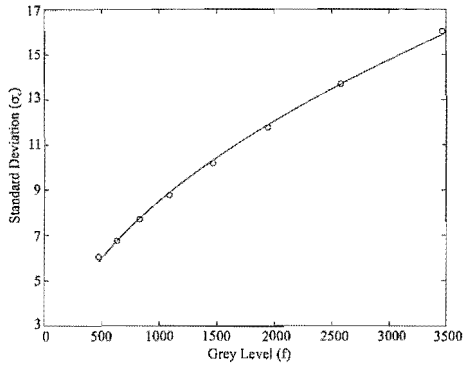


Figure 4.2 Square root model fit $\sigma_c^{mod}(f)$ to the $\sigma_c(k)$ data 'o'.

The dependence of the noise on the grey level f is removed by rescaling the local contrast c_i using the standard deviation of the local contrast $\sigma_c(f)$ as a measure for the noise [150]. Calculation of the local contrast c_i at site i is shown in Equation 3.1 and repeated here for convenience,

$$c_i = f_i - \frac{1}{N}\sum_{j\in\delta_i} f_j, \tag{4.2}$$

where δ_i represents a neighbourhood of pixels centred on i of size N. To obtain $\sigma_c(f)$ the greyscale is divided into non-overlapping but adjacent bins k. A histogram of local contrast values is formed for all pixels with grey levels within the kth bin and this is normalised to form an estimate of the probability density function $p(c|k)$. The value of $\sigma_c(k)$ is then calculated from $p(c|k)$. This is repeated for all greyscale bins.

For direct digital images, where quantum noise is assumed to be dominant, a simple square root relationship is expected between the noise, $\sigma_c(f)$, and the grey level f. Therefore, a square root function fit through the set of points $\sigma_c(k)$ can be used to provide a continuous estimate of $\sigma_c(f)$. This contrast deviation square root model is defined as

$$\sigma_c^{mod}(f) = mf^{1/2} + d, \tag{4.3}$$

where m and d are constants determined by minimising the mean squared error between the model and measured points. This is a variation on the technique described in Chapter 3 which uses cubic spline interpolation to obtain the continuous function $\sigma_c(f)$. Figure 4.2 provides an illustrative example of a square root model fit, $\sigma_c^{mod}(f)$, through a set of $\sigma_c(k)$ points. It is clear from this plot that the square root function provides a good fit to the binned data. A more detailed description of this figure is provided in subsection 4.5.2.

Taking the signal dependency of $\sigma_c^{mod}(f)$ into consideration, the local contrast c_i is equalised by

$$c_i' = \frac{c_i}{\sigma_c^{mod}(f_i)}, \tag{4.4}$$

where c_i' represents the equalised local contrast at site i. It should be noted that in the development of this noise model the mean of the local contrast $\mu_c(f)$ is assumed to be zero. This is to be expected if the noise processes involved have a symmetric distribution, and if the curve relating the grey level to the x-ray exposure is approximately linear within each bin. Also note that f, c, d, $\sigma_c(k)$ and $\sigma_c^{mod}(f)$ are all measured in terms of grey levels, i.e. 0 to 4095 for 12-bit data.

In Chapter 3 a complex binning scheme was required to compensate for the film-screen system's nonlinear response to x-ray photons. Due to the linear detector response of the direct digital modality the specific choice of bins is much less critical. Additionally, the square root model-based approach which is valid for direct digital images, ensures estimation of the relationship between noise variance and intensity remains reliable in those ranges of intensity for which the population of pixels in a particular mammogram is relatively low. This is expected to lead to a more robust estimation of the high frequency image noise, and in turn provide better microcalcification detection results.

The following sections investigate the validity of this square root noise model using both direct digital step wedge phantom images and direct digital mammograms. It should be noted that the filters used and the corresponding window sizes are scaled for images of 100μm per pixel.

4.5 Phantom images

In this section the square root noise model is tested on a set of direct digital step wedge phantom images. Demonstrating that the noise model is accurate for images with no background structure provides initial validation of the assumed quantum noise dominance.

4.5.1 Acquisition

A single step wedge phantom of 10 steps was created by partially overlapping sheets of $5mm$ and $10mm$ thick perspex. Twelve direct digital images were obtained from this phantom using the GE Senographe system at the Nijmegen medical centre. For each image the following image acquisition parameters were varied to represent the range used in routine mammography (see subsection 4.5.4).

Tube Potential Measured in kVp. It is the peak tube voltage from target (anode) to cathode. The *peak* is used because the tube voltage is not constant.

Tube Loading Measured in mAs. This is the product of the current through the x-ray tube (milliampere) and the time the tube was energised (seconds).

Entrance dose Measured in mGy (milligray). This is the dose at the surface of the imaged object.

Target material The target material determines the x-ray spectrum, in specific the characteristic peaks which are used in mammography. Common target materials in mammography are: Molybdenum, Rhodium and Tungsten.

Filter material The filter prevents very low energy x-rays reaching the patient. The penetration of these x-rays in the breast is so low that most x-rays will not reach the detector thus significantly increasing patient dose while only slightly affecting the imaging process. Commonly used filter materials are Molybdenum and Rhodium.

The original 14 bits per pixel GE images were linearly rescaled to 12 bits which is a standard precision used in medical imaging. Each image was 2294×1914 pixels in size and had a 100μm resolution. Figures 4.3(a) and (b) show two example direct digital phantom images. It should be noted that the higher grey levels correspond to the lower attenuation regions in each image.

4.5.2 Noise model application

Implementation of the square root noise model requires c to be binned with respect to grey level f. The obvious method of binning is to place all pixels lying inside one step of the wedge into a single bin, so the number of bins K corresponds to the number of steps in the phantom. This is justified by the fact that pixels associated with a given step are expected to possess similar grey levels. In some phantom images the intensity for one or more of the thinnest and/or thickest steps were in saturation and therefore have been excluded from the calculations. Additionally, because the calculation of local contrast (Equation 4.2) is based on a small neighbourhood, the binning excludes pixels closer to a step boundary than half the width of the neighbourhood.

Figures 4.3(c) and (d) show examples of the local contrast images calculated from the phantom images in Figures 4.3(a) and (b) respectively. The contrast values were obtained using Equation 4.2 with a 9×9 window for $\delta_i(N = 81)$. Justification for this parameter setting

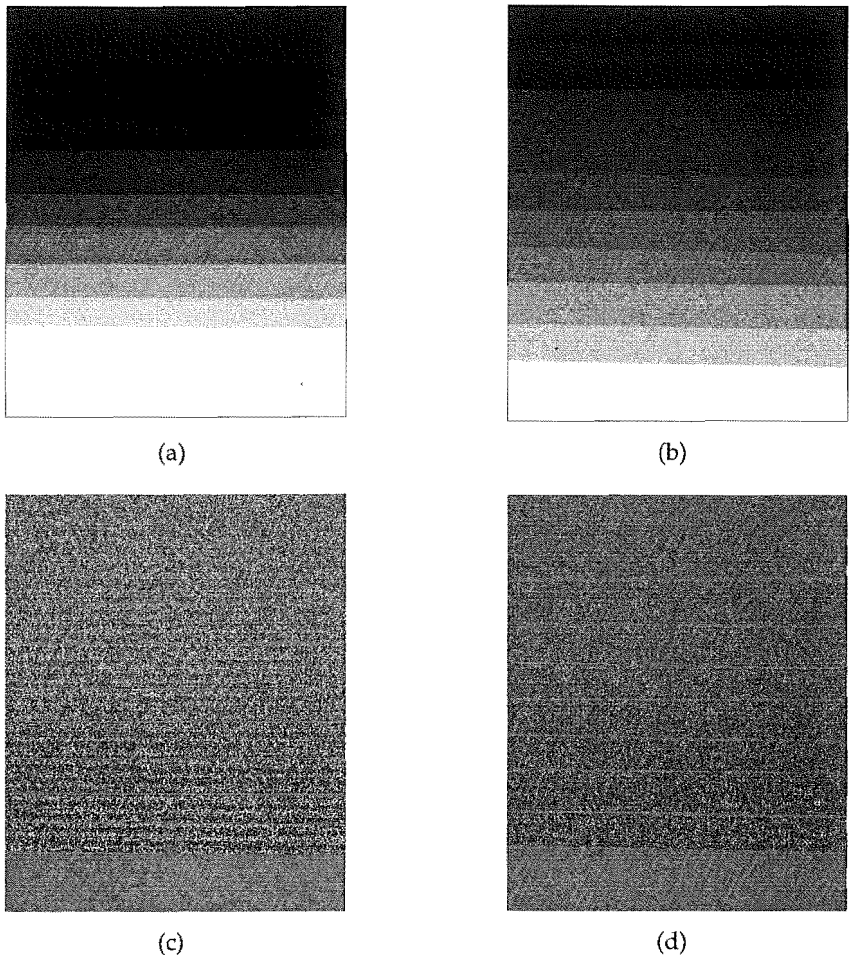


Figure 4.3 (a-b) Example phantom step wedge images. (c-d) Local contrast images calculated using Equation 4.2.

is provided in Section 3.4 and is dependent on the average microcalcification width and average distance between neighbouring microcalcifications expected in practice. Pixels not included in the contrast calculation were set to zero. This includes the portions of the images that were saturated and the step boundary pixels.

The local contrast probability density function $p(c|k)$ is estimated by normalising the histograms of c determined within each bin. Figures 4.4(a) and (b) show examples of four local contrast distributions (corresponding to four different bins) obtained from the local contrast images in Figures 4.3(c) and (d) respectively. Also shown in Figures 4.4(a) and (b) are the corresponding mean grey levels f_k for each bin. These plots illustrate that as the grey level increases, the standard deviation of the local contrast also increases, which is to be expected for a square root noise model. Both contrast plots indicate that $\mu_c(k)$ remains very close to zero.

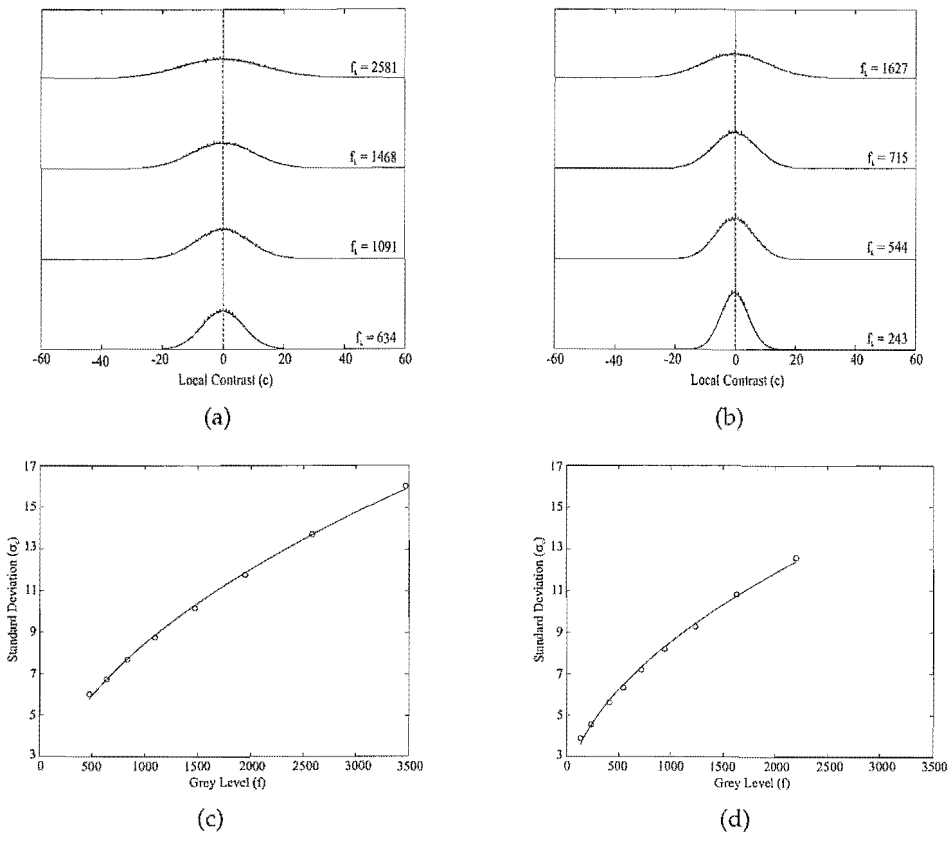


Figure 4.4 (a-b) Plots of $p(c|k)$ for four different bins. (c-d) Contrast standard deviation $\sigma_c^{mod}(f)$ versus grey level. Left hand plots are for the phantom image in Figure 4.3(a), right hand plots are for the phantom image in Figure 4.3(b).

The standard deviation $\sigma_c(k)$ is determined from $p(c|k)$, using

$$\sigma_c(k) = \left(\int_{-\infty}^{\infty} c^2 \, p(c|k) \, dc \right)^{1/2} . \tag{4.5}$$

The best fit continuous function $\sigma_c^{mod}(f)$ is then estimated from Equation 4.3 by minimising the mean squared error between $\sigma_c^{mod}(f_k)$ and $\sigma_c(k)$ for all values of k. Figures 4.4(c) and (d) show the estimates of $\sigma_c(k)$ and the corresponding best fit $\sigma_c^{mod}(f)$. It is clear from these plots that a square root function provides a good fit to the binned data. Quantitative results on the accuracy of the square root noise model are provided in subsection 4.5.4. Once the continuous function $\sigma_c^{mod}(f)$ is obtained, normalisation of the local contrast is calculated using Equation 4.4.

4.5.3 Truncation

The phantom images used in this study were observed to have a small number of artifacts with very high magnitude compared with their surrounding pixels (for example, see Fig-

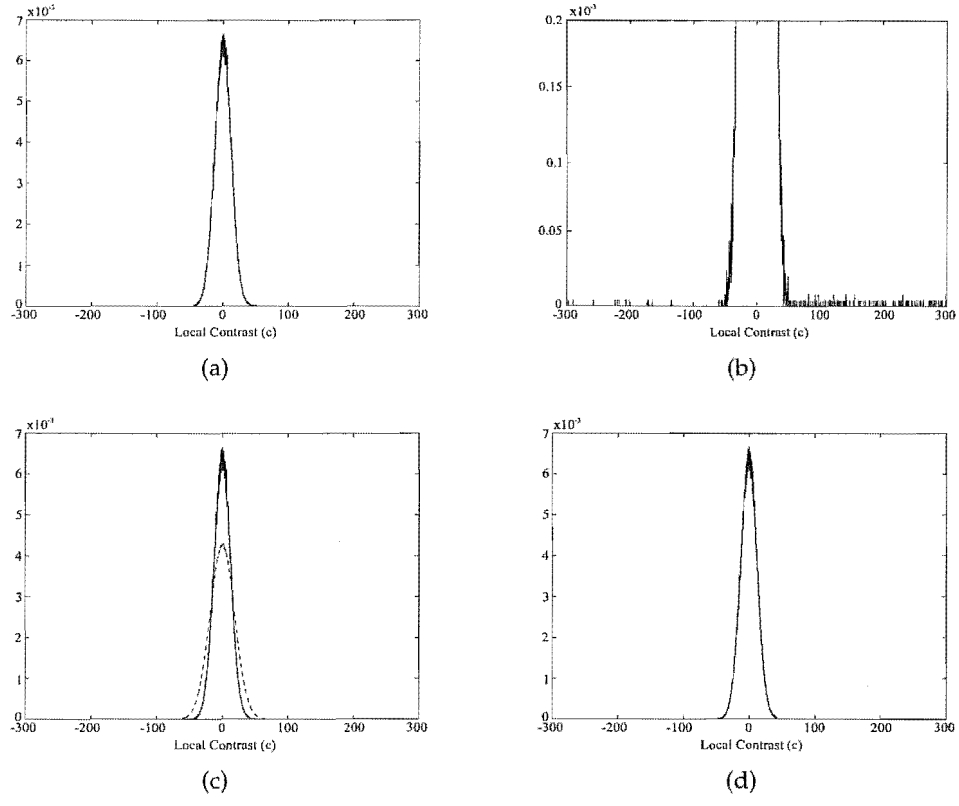


Figure 4.5 (a) Plot of $p(c|k)$ from the phantom image in Figure 4.3(b) for $f_k = 2193$. (b) Closer look at $p(c|k)$ showing the outliers caused by artifact pixels. (c) Gaussian estimation of $p(c|k)$ using the sample standard deviation of the complete data (including outliers). (d) Gaussian estimation using the sample standard deviation of the truncated data with $L = 0.95$. In both cases $\mu_c(k)$ is assumed to be zero.

ure 4.3(b)). The artifacts were determined to be due to imperfections in the material used to make the phantom. The presence of these artifacts cause outliers to exist in the associated probability density function $p(c|k)$, significantly affecting the estimate of $\sigma_c(k)$. Although the likelihood of similar artifacts in actual mammograms is low, a similar situation was noted in Chapter 3 when analysing the noise in film-screen mammograms. In the film-screen case however, the outliers were caused by feature boundary pixels. To ameliorate the problem of the boundary pixels, $p(c|k)$ was truncated to exclude extreme values as described in subsection 3.5.1. Estimates of the noise statistics were then determined based on the truncated distribution [68]. The same method of truncation is used here to remove the problem of artifact pixels.

Figure 4.5 provides an example of how the truncation method works on a single step from the phantom image in Figure 4.3(b). The original $p(c|k)$ is shown in Figure 4.5(a). Figure 4.5(b) provides a closer look at $p(c|k)$ clearly showing the outliers caused by artifact pixels with high magnitude local contrast and low probability. Estimating $\sigma_c(k)$ using the complete local contrast data set, including outliers, results in a value too high for accurate

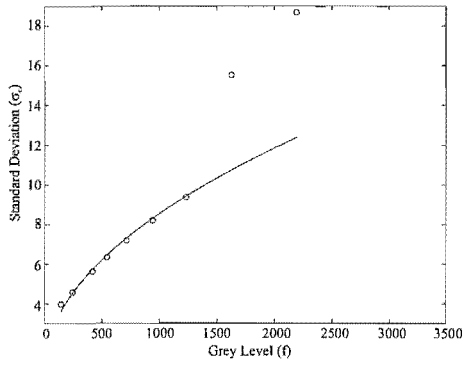


Figure 4.6 Contrast standard deviation versus grey level when truncation is omitted for the phantom image in Figure 4.3(b). The line shows the square root model $\sigma_c^{mod}(f)$ fit to the $\sigma_c(k)$ data for the leftmost 7 points only.

representation of the density function. Figure 4.5(c) provides a visual comparison between $p(c|k)$ and a Gaussian estimation based on $\sigma_c(k)$ of the complete data set. Estimation of $\sigma_c(k)$ based on the truncated data set with $L = 0.95$ (see subsection 3.5.1) provides a much improved representation of the density function and is illustrated in Figure 4.5(d). When calculating $\sigma_c(k)$ in both cases, the value of $\mu_c(k)$ is assumed to be zero.

The noise plots shown in Figure 4.4 were obtained using this method of truncation. Figure 4.6 shows the resulting noise plot for the phantom image in Figure 4.3(b) when estimates of $\sigma_c(k)$ are based on non-truncated data. The probability density functions $p(c|k)$ of the two highest greyscale bins have a significantly higher standard deviation than expected from the square root model. This is due to the artifacts associated with those particular bins. To emphasise the extent to which these two points deviate from the model the square root fit in Figure 4.6 has been estimated using only the leftmost 7 points in the plot. Quantitative results on the improvement in accuracy of the noise estimates when applying the truncation method are provided in the next subsection.

4.5.4 Results

The square root noise equalisation technique was applied to each of the 12 phantom images. Table 4.1 shows the acquisition parameters for each image. As mentioned in subsection 4.5.2, the intensity for one or more of the thinnest and/or thickest steps in some phantom images were clearly in saturation, therefore these steps have been excluded from the calculations. Also shown in Table 4.1 are the noise model parameters m and d from Equation 4.3.

To provide a measure of how well the noise model fits the data the correlation coefficient r between $\sqrt{f_k}$ and $\sigma_c(k)$ was calculated for each phantom image, with a correlation coefficient of 1 representing a perfect fit. For the phantom images the value of r ranged from

Table 4.1 Step wedge phantom image acquisition parameters: Tube Potential (kVp), Tube Loading (mAs), Entrance Dose (mGy), Target Material, Filter Material. The entrance dose was measured by entrance surface air kerma at 5 cm above the bucky. Parameters m and d are obtained from Equation 4.3. The relative mean squared error e determines how well the noise model fits the data.

Image	kVp	mAs	mGy	Target	Filter	m	d	$e(\times 10^{-3})$
1	26	180	14.9	Molybdenum	Molybdenum	0.251	0.423	3.200
2	26	280	23.1	Molybdenum	Molybdenum	0.236	0.734	0.975
3	31	71	8.7	Rhodium	Rhodium	0.251	0.599	0.331
4	31	200	24.4	Rhodium	Rhodium	0.285	-0.079	0.183
5	27	140	11.1	Molybdenum	Rhodium	0.262	0.044	1.100
6	27	100	7.9	Molybdenum	Rhodium	0.250	0.323	1.100
7	30	110	11.2	Molybdenum	Rhodium	0.249	0.493	0.280
8	30	180	18.3	Molybdenum	Rhodium	0.266	0.102	0.170
9	30	180	18.7	Rhodium	Rhodium	0.273	-0.080	0.053
10	30	125	12.9	Rhodium	Rhodium	0.271	-0.077	0.125
11	28	71	7.3	Molybdenum	Molybdenum	0.232	0.666	1.100
12	28	125	12.9	Molybdenum	Molybdenum	0.256	0.192	1.400

0.993 to 0.999 indicating that a square root noise model is an appropriate representation of the image data. This measure however does not take into account any linear scaling of $\sigma_c^{mod}(f)$, therefore an additional accuracy measure, the relative mean squared error e, was calculated as

$$e = \frac{\sum_{k=1}^{K} (\sigma_c(k) - \sigma_c^{mod}(f_k))^2}{\sum_{k=1}^{K} (\sigma_c^{mod}(f_k))^2}. \tag{4.6}$$

A relative mean squared error of 0 indicates a perfect fit to the data. The maximum value of e for all 12 phantom images is 3.2×10^{-3}. This small error value confirms that $\sigma_c^{mod}(f)$ is an accurate measure of the high frequency noise.

The noise model values and accuracy measurements shown in Table 4.1 were all obtained using the truncated distribution method with $L = 0.95$. When the values of m and d were obtained without truncation the relative mean squared error increased in some cases by 2 orders of magnitude, indicating that truncation can significantly improve the accuracy of the noise estimation.

In a strictly quantum limited system the noise present at zero exposure (zero grey level) would be zero. This is represented in the noise model by the parameter d. Table 4.1 shows that the estimated values for d vary slightly around zero for different image acquisition conditions. In order to determine a representative noise value for zero exposure, an image

Table 4.2 Parameter m obtained from Equation 4.3, and the relative mean squared error e when fixing $d = 0.633$.

Image	m	$e(\times 10^{-3})$
1	0.244	3.400
2	0.239	1.000
3	0.250	0.333
4	0.269	0.532
5	0.245	1.900
6	0.240	1.400
7	0.245	0.311
8	0.253	0.439
9	0.257	0.373
10	0.254	0.544
11	0.233	1.100
12	0.243	2.000

from the GE Senographe system with a $4mm$ aluminium plate placed in front of the detector was obtained. The contrast standard deviation $\sigma_c(0)$ was calculated to be 0.633, which is similar in magnitude to the estimated d values. This very small positive value could be due to a small additional noise source present within the direct digital imaging system.

When the remaining noise model parameter m was recalculated for the 12 phantom images with d fixed at 0.633, the value of the error e necessarily increased. Table 4.2 provides the values of both m and e for the fixed y-intercept. The maximum value of e for all images in the data set is 3.4×10^{-3}. This is still an acceptably small error value, suggesting that the noise model estimation could be reduced to finding a single parameter. Chapter 8 provides a comparison in microcalcification detection performance for both the 1-parameter and 2-parameter methods. These results indicate that fixing parameter d slightly reduces the sensitivity of the CAD scheme, therefore in the remainder of this chapter the 2-parameter method is used.

4.6 Direct digital mammograms

In this section the square root noise model is applied to a set of direct digital mammogram images. Results comparable with those for the phantom images are obtained, although differences in noise characteristics are observed near the periphery of the breast.

4.6.1 Acquisition

The direct digital mammogram data set used in this application is described in detail in Section 2.2. It consists of 124 mammograms composed of both cranio-caudal and medio-lateral oblique views from 33 different patients.

4.6.2 Noise model application

For the phantom images in subsection 4.5.2, pixels from each step in the wedge were placed into a single bin. When implementing the noise model on direct digital mammogram images, however, binning cannot be performed in the same manner. To obtain $\sigma_c(k)$, the greyscale is divided into non-overlapping adjacent bins numbered $k = 1, 2, \ldots, K$, with each bin built up of an approximately equal number of pixels (see the fixed size binning scheme described in subsection 3.5.2). For the purposes of comparison with the noise equalisation method for the film-screen modality, the number of bins K is set equal to 16. Figures 4.7(a) and (b) show two raw mammogram images from the direct digital data set and Figures 4.7(c) and (d) illustrate the corresponding local contrast images calculated using Equation 4.2. The contrast values were obtained using a 9×9 window for $\delta_i(N = 81)$. Only pixels associated with the breast tissue area of each mammogram are included in the contrast calculation, all other pixels are set to zero (see Section 2.3).

The local contrast probability density function $p(c|k)$ is estimated by normalising the histograms of c determined within each bin. Figures 4.8(a) and (b) show examples of four local contrast distributions (corresponding to four different bins) obtained from the local contrast images in Figures 4.7(c) and (d) respectively. Also shown in these plots are the corresponding mean grey levels f_k for each bin. Once again the contrast plots indicate that as the grey level increases, the standard deviation of the local contrast also increases, and $\mu_c(k)$ remains very close to zero.

The standard deviation $\sigma_c(k)$ is calculated from $p(c|k)$ using the truncated distribution method described in subsection 3.5.1. The best fit continuous function $\sigma_c^{mod}(f)$ is determined using Equation 4.3 and the local contrast feature c_i is equalised by Equation 4.4 with $\mu_c(f)$ set to zero for all f.

Although the phantom image results are in agreement with the theory, it was found that in direct digital mammogram images there is a significant deviation from the square root noise model in the peripheral area of the breast, where the tissue is uncompressed. Pixels associated with this uncompressed breast area are easily identified using the routinely recorded breast tissue thickness of each mammogram image. Figure 4.9 shows that a simple semi-circular model can be used to represent the uncompressed breast area. On a mam-

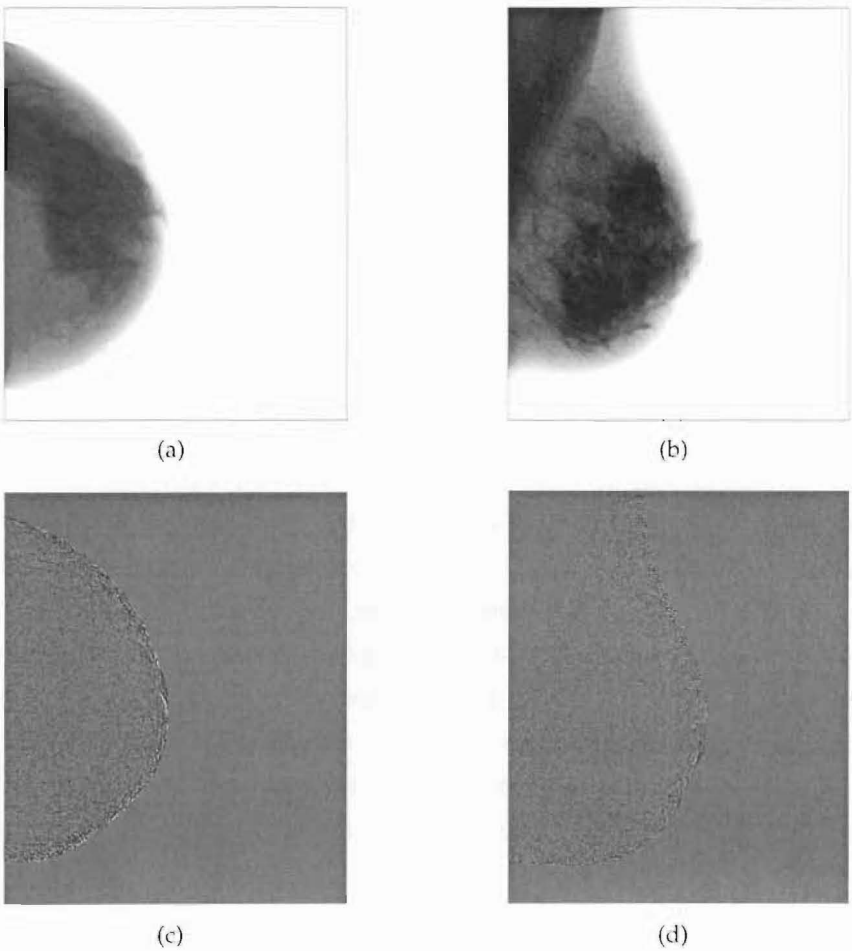


Figure 4.7 (a-b) Example direct digital mammograms. (c-d) Local contrast images calculated using Equation 4.2.

mogram this region is recognised as the area from the skin line moving in a distance of half the recorded breast tissue thickness.

Figures 4.8(c) and (d) show the estimates of $\sigma_c(k)$ and the corresponding best fit $\sigma_c^{mod}(f)$ obtained when omitting the uncompressed area of the breast. The plots indicate that with the exclusion of the peripheral breast area a square root noise model is a good representation of the noise dependency on the grey level. When all pixels within the breast are included in the model a square root noise plot such as that shown in Figure 4.10 is obtained. The probability density functions $p(c|k)$ of the two highest greyscale bins have a significantly higher standard deviation than expected from the square root model. These bins are associated with the uncompressed area of the breast.

Recent experiments performed by researchers in Nijmegen show that this deviation from the expected noise model is due to irregularity of the skin surface, which is amplified near

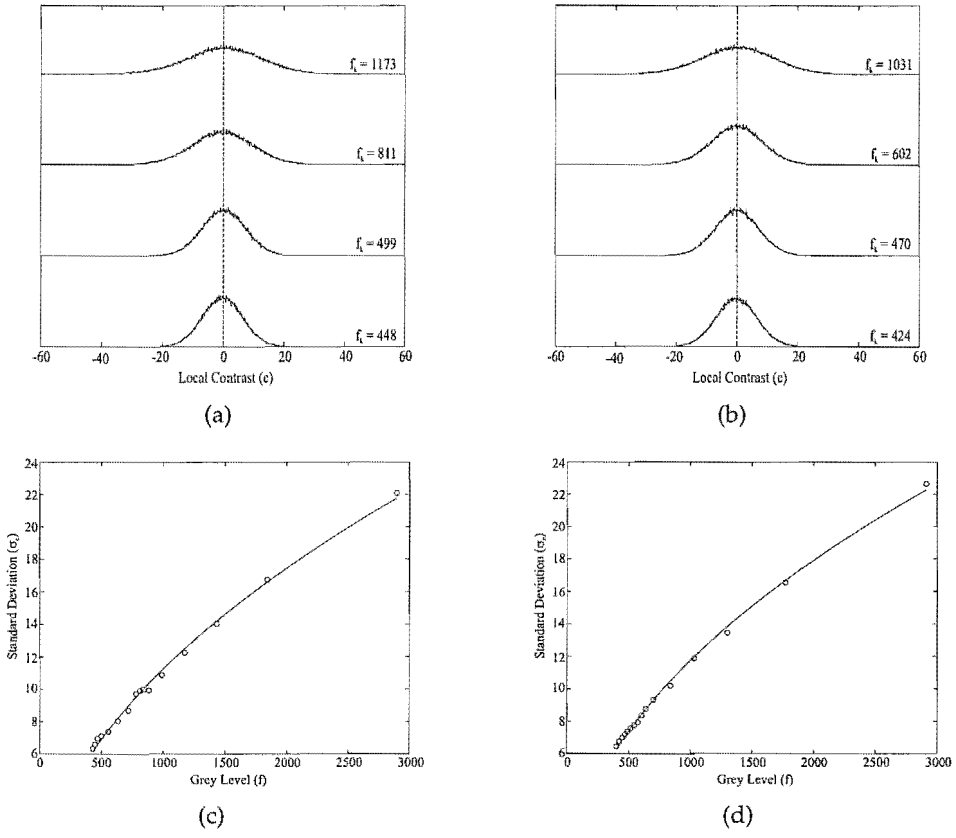


Figure 4.8 (a-b) Plots of $p(c|k)$ for four different bins. (c-d) Contrast standard deviation $\sigma_c^{mod}(f)$ versus grey level. Left hand plots are for the mammogram in Figure 4.7(a), right hand plots are for the mammogram in Figure 4.7(b).

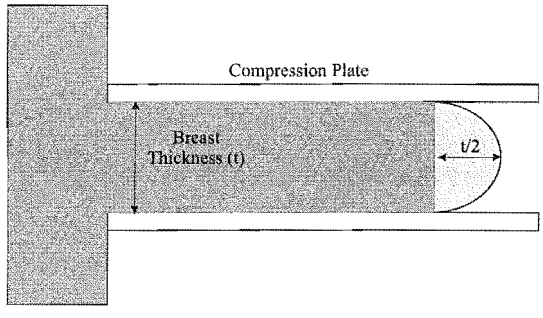


Figure 4.9 Illustrating how the uncompressed breast area (lightly shaded area) is determined from the breast thickness (t).

the breast edge due to the small angle between the skin surface and the incident x-ray beam [74]. Compensation for this effect has not been included as part of this research, therefore in all the results hereafter the breast area of each direct digital mammogram is eroded from the skin line to exclude the uncompressed area of the breast as explained in the Figure 4.9.

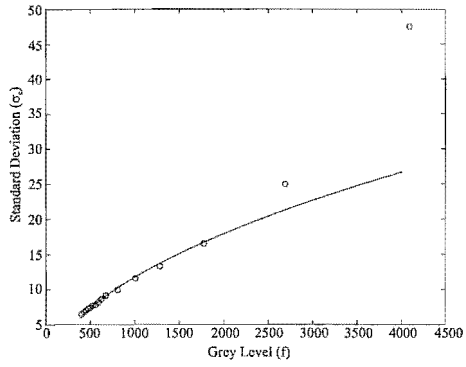


Figure 4.10 Contrast standard deviation versus grey level when including pixels from the uncompressed area of the breast for the mammogram in Figure 4.7(b). The line shows the square root model $\sigma_c^{mod}(f)$ fit to the $\sigma_c(k)$ data for the leftmost 14 points only.

4.6.3 Results

The square root noise equalisation technique was applied to each of the 124 direct digital mammogram images. As for the phantom cases two measures were used to determine how well the noise model fit the data: the correlation coefficient r and the relative mean squared error e. The correlation coefficients for the two mammogram images shown in Figures 4.7(a) and (b) were 0.995 and 0.996 respectively. For all images in the direct digital data set the value of r ranged from 0.800 to 0.998, with 70% having a value of r greater than 0.950. This indicates that the square root noise model is an appropriate representation of the image data, however, once again r does not take into account any linear scaling of $\sigma_c^{mod}(f)$.

The relative mean squared error e for the mammograms in Figures 4.7(a) and (b) were 4.836×10^{-4} and 3.093×10^{-4} respectively. For all images in the data set, e ranged from 2.256×10^{-4} to 4×10^{-3}, with 70% having a value less than 2×10^{-3}. These small error values confirm that the square root noise model is accurate for application to direct digital mammogram images.

Chapter 8 provides a comparison between the microcalcification CAD results when using the noise model developed in Chapter 3 and the square root noise model developed here. These two techniques are also compared with the results obtained when noise equalisation is omitted. The results provide clear evidence that high frequency noise equalisation is important for the detection of microcalcifications. It also shows that the use of the square root model leads to a more robust estimation of the high frequency image noise in the direct digital modality, and thus better microcalcification detection.

4.7 Equalisation examples

This section provides examples illustrating the results obtained after application of the square root noise equalisation method. Figures 4.11(a) and (b) show two sub-images, 480 × 480 pixels in size, obtained from two direct digital mammograms, both containing microcalcification clusters. Note that both images are raw direct digital images. Calculation of the local contrast using Equation 4.2 results in the images shown in Figures 4.11(c) and (d). These images highlight the microcalcifications present by producing extreme negative local contrast values at their locations. Application of the noise equalisation method results in an equalised local contrast image (calculated using Equation 4.4) shown in Figures 4.11(e) and (f).

The local contrast images illustrate how equalisation of the image noise provides clearer visualisation of the microcalcifications present. In a CAD system, this ensures that separation of microcalcifications from background mammographic structure is easier and detection sensitivity improved. It is important to note that care is needed to make sure the system manufacturer has not applied some form of postprocessing to the mammograms which may invalidate the model; *raw* images should be used.

As mentioned in Chapter 3 the equalised local contrast images not only highlight microcalcifications but also curvilinear structures (locally linear features). Detection of these structures within a mammogram image is described in detail in Chapter 5.

4.8 Summary

This chapter provided detailed experimental results for a novel noise equalisation method specific to direct digital mammogram images. It is the first known report discussing the detection of microcalcifications in direct digital mammograms using a theoretical model for the dependency of noise variance on intensity. The model allows a straightforward and accurate approach to noise equalisation leading to a robust estimation of the image noise when compared to the film-screen modality.

Section 4.4 described an accurate adaptive approach for modelling the image noise utilising the known quantum noise dominance. In this method a high pass filtered version of the original image, known as the local contrast, was used to represent the high frequency noise contribution associated with each image. Utilising the linear detector response of a direct digital system provides a simple square root relationship between the noise and the grey level. This information was used to equalise the local contrast, removing the signal dependency of the noise.

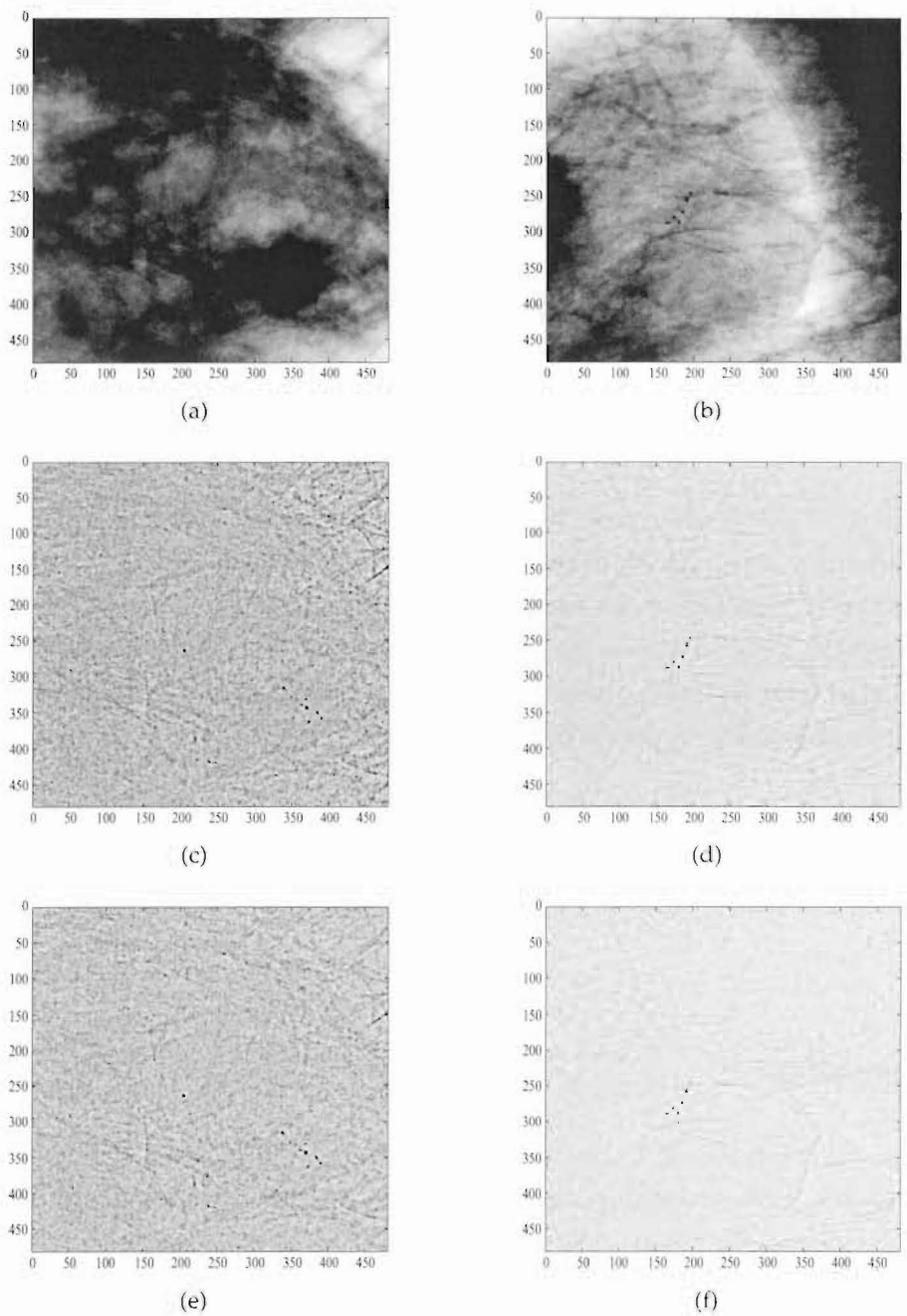


Figure 4.11 (a-b) Two sub-images from two raw direct digital mammograms, both containing microcalcification clusters. (c-d) Local contrast image. (e-f) Equalised local contrast image.

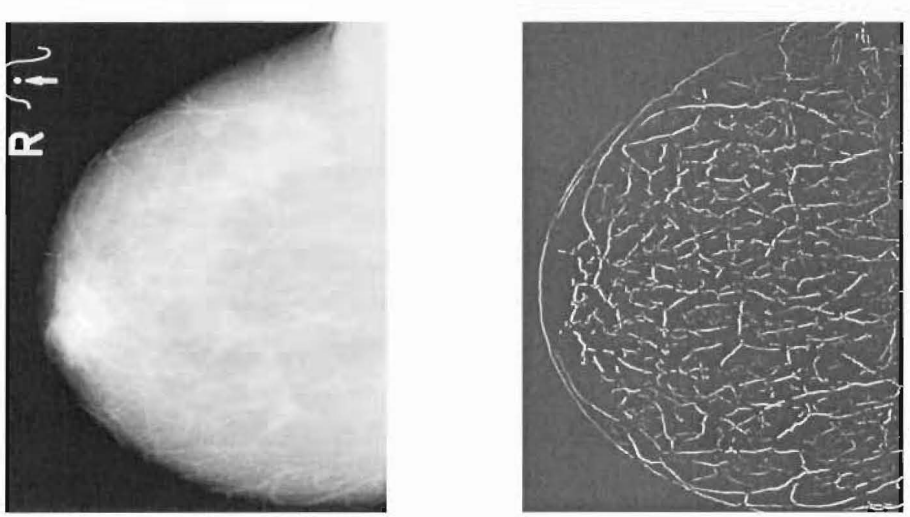
Sections 4.5 and 4.6 investigated the validity of the square root noise model using a set of 12 step wedge phantom images and 124 direct digital mammogram images. Two measures were used to quantify the accuracy of the model (correlation coefficient and the relative mean squared error). It was found that artifacts within the phantom images cause overestimation of the noise. This estimation is improved by truncating the local contrast distributions to exclude outliers providing a more reliable and accurate representation of

Chapter 5

Shape parameter for curvilinear structure detection

5.1 Introduction

Curvilinear structures (CLS) are locally linear high intensity structures in mammograms with a well-defined local orientation that varies along the length of the structure. Figure 5.1 provides an example of a digitised film-screen mammogram and a corresponding image displaying the pixels identified as belonging to a CLS. Note that in Figure 5.1(b) high intensity pixels indicate a high probability of CLS.



(a) (b)

Figure 5.1 (a) A digitised film-screen mammogram. (b) Image highlighting the CLS present.

the noise level. Although the likelihood of similar artifacts in actual mammograms is low, the same method gives good results when applied to the direct digital mammograms.

Further improvement in the noise model occurred when excluding pixels associated with the uncompressed area of the breast. There is a significant deviation from the square root model in this area due to noise processes other than quantum noise playing a significant role.

A slight alteration in the basic square root model was tested by fixing d, the residual contrast standard deviation at zero exposure. This simplifies the estimation process to finding a single parameter however causes a slight decrease in the noise model accuracy.

Overall, the small error values reported in this chapter confirm that the square root noise model is accurate for application to direct digital images.

CLS correspond physically to either milk ducts, blood vessels, Cooper's ligaments or spicules in the breast. With the exception of spicules, all of these structures comprise normal healthy breast tissue. The presence of CLS complicate, and often defeat, image analysis tasks such as the detection of microcalcifications [61, pp. 225-231]. This is because the local operations that are used to detect microcalcifications often respond strongly to CLS (see Figures 3.13 and 4.11). When detecting microcalcification clusters it is important that the appearance of both microcalcifications and CLS are modelled to ensure accurate and robust CAD results.

A distinction between faint microcalcifications and CLS can only be made on the basis of shape, as these two structures have similar contrast. This chapter describes the development of a new shape parameter for identifying CLS which utilises and extends a method recently proposed by Kovesi [86]. The method treats features in terms of their Fourier components and is know as *phase congruency*.

Section 5.2 gives a review of the line-like detection methods previously developed in the literature. One such method specifically designed for the detection of CLS is detailed in Section 5.3. This method, proposed by Karssemeijer [71], is calculated using an algorithm resembling the Hough transform. Identifying the limitations of the technique provides motivation for the development of a new shape parameter using a multiresolution oriented approach. This is provided in Section 5.4 with the development of a new shape parameter based on the method of phase congruency. Section 5.5 discusses the advantages and disadvantages of both the Hough and phase congruency shape parameters, concluding that the phase congruency shape parameter is superior in coping with the varying size and orientation of CLS and is less computationally expensive.

5.2 Existing methods

From an image processing point of view, the detection of CLS appears to correspond to the general problem of ridge finding, however standard ridge detection methods do not work well on mammographic images since CLS do not conform to a simple ridge model. A major problem stems from the fact that CLS appear at different depths within the breast, that is, they appear at an unpredictable range of magnitudes, scales and orientations. A shape parameter designed specifically to highlight CLS must be capable of adapting to the varying properties of these features. Additionally, in this application a distinction between faint microcalcifications and CLS is crucial. Therefore, not only must the shape parameter be capable of highlighting CLS, it must also suppress microcalcifications.

The simplest and most common method of line-like structure detection is the Hough transform [37, 71, 83, 139]. Traditionally this transform has been used to detect edges within

an image, however simple modifications can be made to ensure the detection of line-like structures. Karssemeijer developed a modified Hough transform method using the local probability density function of gradient directions to determine the presence or absence of local line-like features [71]. The method is applied within a window of fixed size moved over the entire image. A detailed description of this method is given in the next section and is referred to as the Hough shape parameter, although it is not strictly a Hough transform method. The technique is designed specifically to highlight line-like features while simultaneously suppressing blob-like structures and is compared with the new shape parameter described in Section 5.4. The disadvantages of this Hough shape parameter include a high execution time and cumbersome implementation [102, 103]. Additionally, the varying size of CLS within a mammogram image means the success of this method is highly sensitive to the size of the moving window.

Other researchers have used texture-based approaches in detecting line-like structures within the breast [80, 81]. Kegelmeyer *et al* implements an approach based on statistical analysis of a map of pixel orientations for the detection of spicules associated with stellate masses [81]. The idea is that if an excess of pixels orientated towards a given region is found, then this region may be suspicious. Laws textural energy features are used in conjunction with this measure to represent the range of possible manifestations of normal breast tissue [7, 122, 155], thus reducing the number of false positive spiculated mass detections. It is noted that no attempt is being made to identify spicules (line-like structures) explicitly, thus making it inappropriate for CLS detection.

The variable size and orientation of CLS suggest the implementation of a multiresolution approach. Karssemeijer and te Brake developed a well known line-like feature detector using a multiscale implementation [77]. At a given scale, accurate line-based orientation estimates are obtained from the output of three-directional, second-order, Gaussian derivative operators. The orientation at the scale at which these operators give a maximum response is selected. If a linear structure is present at a given site, this method provides an estimate of the orientation of this structure, whereas in other cases the image noise generates a random orientation. The pixel orientations are then used to construct an operator which is sensitive to straight lines, thus detecting the line-like features within the image. However, again the method was developed specifically for the detection of spiculated lesions and in its current form does not provide a good distinction between CLS and microcalcifications.

Other multiresolution oriented line-like feature detection techniques use various wavelet filter banks, the most common of which is known as the steerable filter [46, 84, 133, 137]. The term *steerable filter* is used to describe a class of filters in which a filter of arbitrary orientation is synthesised as a linear combination of a set of *basis filters*. These particular filters are capable of picking up features of all types (i.e. lines and edges). Local phase

information is used to determine the type of feature present. It should be noted that aspects of this wavelet filter bank method are similar to the new shape parameter presented in this chapter.

5.3 Hough shape parameter

The shape parameter developed by Karssemeijer [71] for the detection of CLS has been singled out for further discussion in this section. The parameter was specifically designed to aid in the detection of microcalcifications making it very applicable to this study.

Karssemeijer's method resembles the Hough transform, which is used in image analysis for the detection of straight lines [37]. The shape parameter ζ_i is extracted from the image data in a small neighbourhood of i using a window of fixed size moved over the entire image. This shape parameter indicates whether a line-like or blob-like feature at site i is likely. The method utilises the gradient magnitude η_i and orientation φ_i of pixels within each neighbourhood which are accumulated in an (η, φ) table (accumulator array). Calculation of ζ_i from the local data is described extensively in [71] and is explained in the following with respect to the examples in Figure 5.2.

Figures 5.2(a), (c) and (e) show binary simulated images of three features possible within a small neighbourhood. For the purpose of this example the size of the moving window (which corresponds to size of the neighbourhood) has been set to 50×50 pixels. At each site in the small neighbourhood the orientation φ_i and magnitude η_i of the gradient are calculated using a 7×7 Sobel operator. The gradient magnitude is normalised to range from $0 - 1$ by dividing by the maximum gradient magnitude within the entire image and the orientation is calculated in the interval $0 - 2\pi$. Figures 5.2(b), (d) and (f) show the accumulator arrays obtained when applying the Sobel operator to the images in Figures 5.2(a), (c) and (e) respectively. Higher grey level values correspond to a higher number of pixels with a certain magnitude and orientation. Because of the binary nature of the examples, a large number of pixels have zero gradient magnitude and an undefined orientation. These pixels have been ignored for the purposes of this illustration.

It appears from Figure 5.2 that a distinction between different feature types can be determined solely on the distribution of gradient orientations. The accumulator array for the blob-like feature possesses a very broad distribution of orientations. For the line-like feature however there are two dominant orientations at approximately $\pi/2$ and $3\pi/2$. Both these orientations are perpendicular to the major axis of the line feature. In the case of the step feature only one dominant orientation is observed.

The probability density function of gradient orientations $p_i(\varphi)$ is estimated by summing the values in the accumulator array at each orientation for $\eta \geq T$, with T a fixed threshold

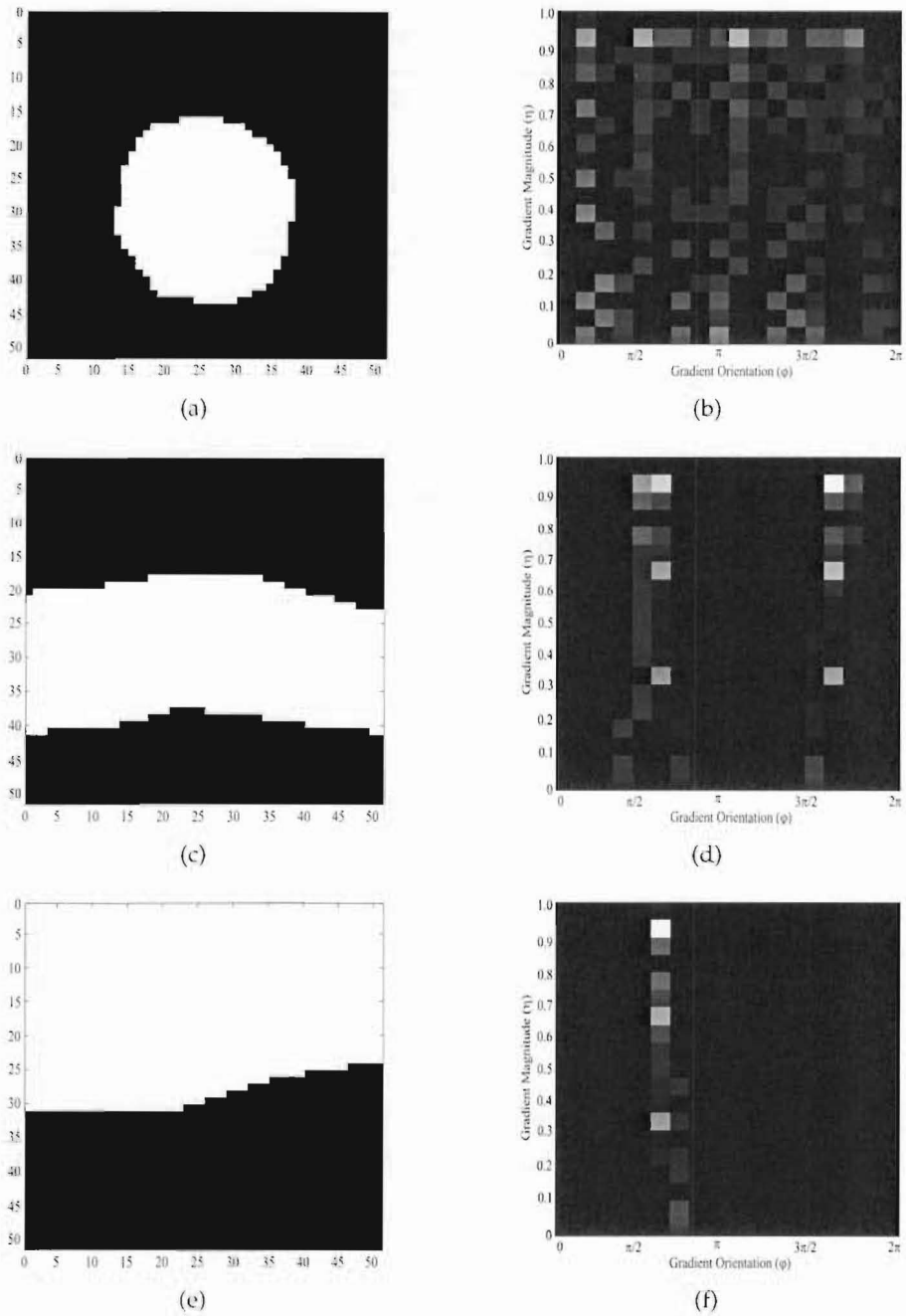


Figure 5.2 Illustration of how the accumulator array for the Hough shape parameter is determined from three different feature types. (a,b) Binary blob-like feature and corresponding accumulator array. (c,d) Binary line feature and corresponding accumulator array. (e,f) Binary step edge and corresponding accumulator array.

applied to reduce the influence of noise. In this example and in other results to follow the value of T is set such that half the pixels in the small neighbourhood have a gradient magnitude greater than T.

Table 5.1 Hough shape parameter ζ_i calculated for the three different features types in Figure 5.2.

Feature	Shape Parameter (ζ_i)
Blob	1.4203
Line	0.4531
Step	0.8412

To normalise the φ-scale a coordinate transform is performed by firstly translating the least occurring gradient direction to $\varphi = 0$. The mean gradient direction $\overline{\varphi}$ is then computed in the interval $[0, 2\pi)$ and an additional translation is carried out to put $\overline{\varphi}$ at $\varphi = \pi$. This normalised parameter space is used to calculate the shape parameter as

$$\zeta_i = \sqrt{\int_0^{2\pi} (\varphi - \varphi_b)^2 p_i(\varphi) d\varphi}, \tag{5.1}$$

where $\varphi_b = \frac{1}{2}\pi$ for $\varphi \in [0, \pi)$ and $\varphi_b = \frac{3}{2}\pi$ for $\varphi \in [\pi, 2\pi)$. Low values of ζ_i indicate that a line-like structure at i is likely.

Table 5.1 provides the Hough shape parameter values calculated for the features in Figure 5.2. These correspond to the parameter values for the centre pixel of each neighbourhood. As can be seen the lowest shape parameter value corresponds to the line-like feature while the highest value corresponds to the blob-like feature indicating that ζ_i is capable of distinguishing between these two feature types.

Unfortunately, this Hough shape parameter method suffers from a number of drawbacks. First, the success of the Hough method is highly dependent on the size of the moving window. The window must be chosen large enough to encompass each individual feature while at the same time be small enough to ensure multiple features are not covered simultaneously. Second, the fixed Hough window size does not allow varying feature sizes to be accurately represented which is an important property of CLS. Third, the method is computationally expensive. Each of these limitations are discussed in detail in Section 5.5 when applying the Hough shape parameter to both a simulated binary image and a digitised film-screen mammogram from FSD-B.

Identification of the drawbacks associated with the Hough shape parameter suggests the need for a multiresolution approach that is not highly dependent on the parameters chosen. With this in mind a new shape parameter was designed based on the method of phase congruency.

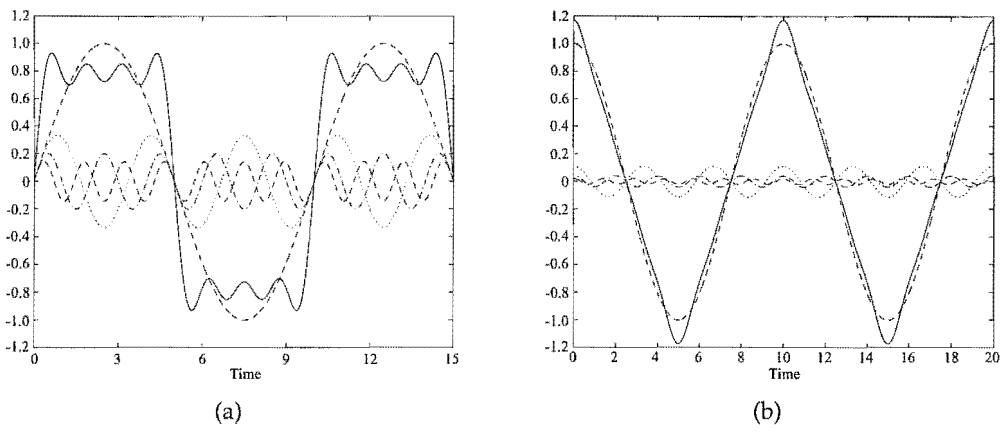


Figure 5.3 Construction of a (a) square wave and (b) triangular wave from their Fourier series. The first few terms of the respective Fourier series are plotted with broken lines, the sum of these terms is the solid line. The Fourier components are all in phase at the zero crossings of the step in the square wave, and at the peaks and troughs in the triangular wave.

5.4 Phase congruency shape parameter

Phase congruency was first proposed by Morrone *et al* [111, 112] and Kovesi [86] as a computational model for the perception of low-level features such as step edges and lines in images. Representation of the image is given in terms of the local variation of amplitude and phase with frequency, that is, the method treats features in terms of their Fourier components rather than in terms of intensity gradients.

Phase congruency utilises the fact that features are perceived at points in an image where the Fourier components are maximally in phase. For example, Figure 5.3(a) shows the first few terms of the Fourier series that represents a square wave. All the Fourier components are sine waves that are exactly in phase (i.e. congruent) at every zero crossing of the square wave. The phase angle at which congruency occurs is 0° or 180° depending on whether the step is positive or negative, respectively. At all other points in the square wave individual phase values vary, making phase congruency lower. Similarly observation of Figure 5.3(b) indicates that for a triangular waveform phase congruency is a maximum at both the peaks (90°) and troughs (270°).

In calculating phase congruency it is important to obtain spatially localised frequency information in images. However, a point of phase congruency is only considered significant if it occurs over a wide range of frequencies. To satisfy both these requirements many scales must be considered simultaneously, implying a multiresolution approach. Additional evidence supporting the multiresolution approach stems from the detection of CLS of varying size. This requires the same broad frequency range and multiple scales. As shown in Fig-

ure 5.3 different feature types give rise to points of phase congruency and it is the phase angle at which the congruency occurs that distinguishes which feature is present.

The following subsections describe in detail the calculation of a new phase congruency shape parameter using a multiresolution wavelet filter bank. The filters are first described in one dimension and then extended to two dimensions for application to a mammogram image. The shape parameter is determined based on the rich set of local property measures from the phase congruency approach. This parameter is designed to highlight line-like features (CLS) while suppressing blob-like structures (microcalcifications).

5.4.1 Calculating phase congruency via wavelets

Wavelets offer perhaps the best method of obtaining spatially localised frequency information in images. The wavelet transform is used to obtain frequency information local to a point an image. To ensure that phase is preserved linear-phase filters are used. These are nonorthogonal wavelets in symmetric/antisymetric pairs. Using two filters in quadrature enables calculation of the amplitude and phase of a signal for a particular frequency at a given spatial location.

The objective of the wavelet design is to obtain a wide range of frequency information with maximal spatial localisation. Here the approach suggested by Morlet *et al* [110] is followed. However, rather than using Gabor filters, logarithmic Gabor filters are preferred as suggested by Field [43]. Log Gabor filters have a Gaussian transfer function in the logarithmic frequency domain. They allow filters with arbitrarily large bandwidth to be constructed while still maintaining a zero DC component in the even-symmetric filter. On the linear frequency scale, the log Gabor filter has a transfer function of the form

$$G(\omega) = \exp\left[\frac{-(\log(\omega/\omega_o))^2}{2(\log(\kappa/\omega_o))^2}\right], \tag{5.2}$$

where ω_o is the filter's centre frequency. To ensure all successive filters are related to each other by a fixed scaling constant (constant shape ratio). The term κ/ω_o must be held constant for varying ω_o, where κ is the log Gaussian standard deviation of the filter. For example, a κ/ω_o value of 0.75 results in a filter bandwidth of approximately one octave and a value of 0.55 results in a two-octave bandwidth. Figure 5.4 shows three log Gabor filters of different bandwidths all tuned to the same centre frequency.

Since the size of each feature to be detected is unknown, the design of the wavelet filter bank must be such that the combination forms an even coverage of an appropriate range of the frequency spectrum. Uneven coverage of the spectrum would cause features at some scales to be considered less prominent than others simply because of the uneven filter coverage.

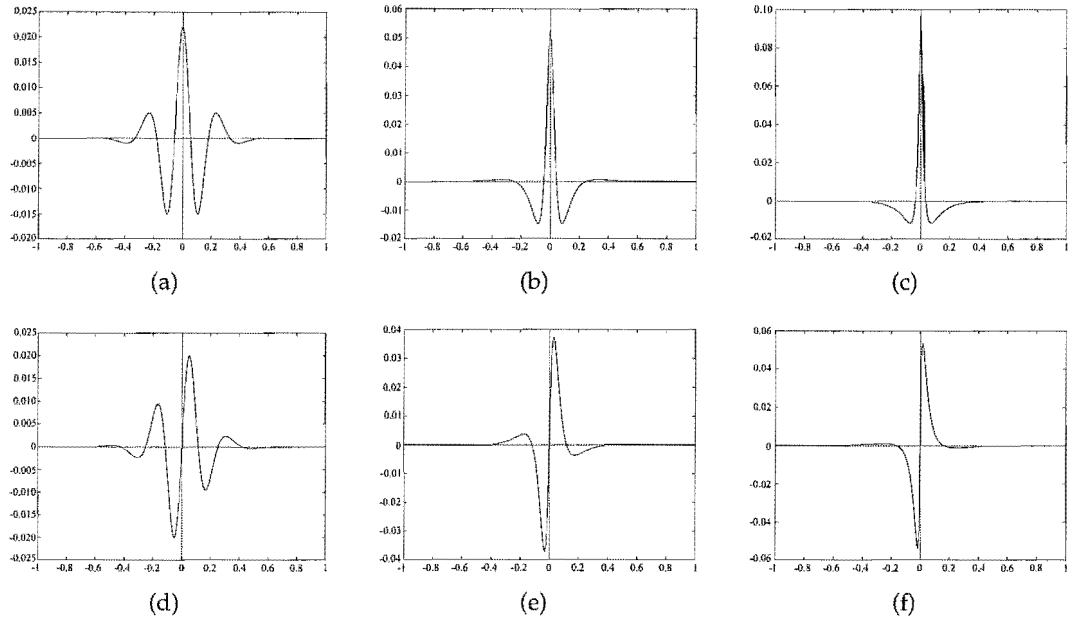


Figure 5.4 Three quadrature log Gabor wavelets all tuned to the same frequency, but having bandwidths of 1 (left), 2 (middle), and 3 (right) octaves respectively. The top plots represent the even-symmetric log Gabor filters, while the bottom plots represent the odd-symmetric filters.

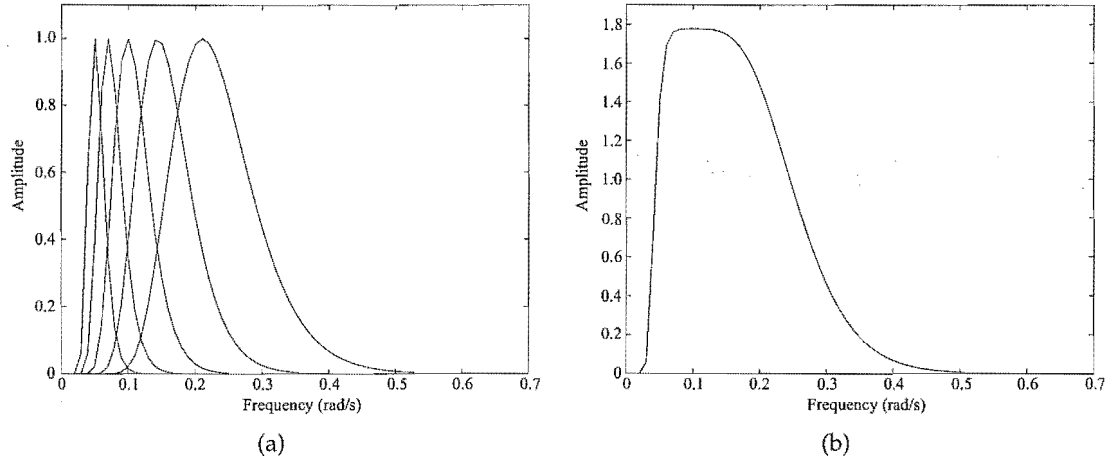


Figure 5.5 (a) Log Gabor transfer functions of five wavelet filters. (b) Sum of the 5 wavelet filter transfer functions producing an even coverage of the frequency spectrum.

In designing the wavelet filter bank a compromise must be found between an even coverage of the frequency spectra while at the same time minimising the number of filters used so as to minimise the computational requirements. Even-symmetric log Gabor filters, by definition (see Equation 5.2), have a zero DC component irrespective of the filter's bandwidth. This allows arbitrarily large bandwidth filters to be constructed, thus keeping the number of filters required for even coverage to a minimum. Figure 5.5(a) shows an example of a wavelet filter bank consisting of 5 wavelet filters each covering different regions of the frequency spectrum. Figure 5.5(b) shows the sum of the individual filters in

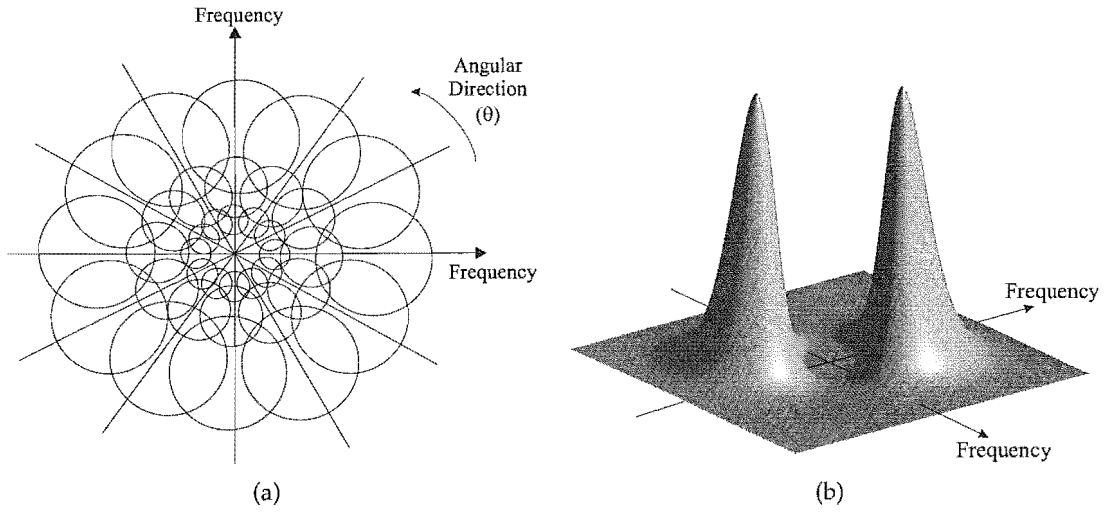


Figure 5.6 (a) Tiling of the 2-D frequency plane with oriented filters at different scales. Note that the DC frequency value is at the centre of the plane. (b) 3-D plot of an even-symmetric filter transfer function at a single scale and orientation.

Figure 5.5(a) producing an approximately uniform amplitude for the band of frequencies covered by the filters.

So far this discussion has been limited to signals in one dimension. Calculation of phase congruency requires a 90° phase shift of the signal, which is done using odd-symmetric filters. Since, rotationally symmetric odd-symmetric filters cannot be constructed, the analysis of a two dimensional image is achieved by applying the one-dimensional analysis over several orientations and combining the results. Three design issues must be resolved when changing to 2-D: the shape of the filters in two dimensions, the number of orientations to analyse, and the way in which the results from each orientation are combined. The first two issues are dealt with in this subsection, while the last is left till subsection 5.4.2.

When extending to two dimensions it is important to ensure that there is no corruption of the phase data in the image. A one-dimensional filter can be extended to two dimensions by simply applying a spreading function across the filter perpendicular to its orientation. Such a 2-D filter is separable allowing image convolution to be accomplished by a 1-D convolution with the spreading function, followed by a 1-D convolution in the orthogonal direction with the wavelet function. The obvious spreading function to use is the Gaussian. Any function smoothed with a Gaussian undergoes amplitude modulation of its components, but phase is unaffected. This ensures the phase congruency at any feature is preserved.

To detect features at all orientations, the bank of filters must be designed so that it tiles the frequency plane uniformly (See Figure 5.6(a)). In the radial direction, within the frequency plane, the filters are designed in the same way as the design of the 1-D filter bank (that is,

log Gaussians with geometrically increasing centre frequencies and bandwidths). Thus the filters appear as 2-D functions symmetrically or anti-symmetrically placed with respect to the origin, depending on their spatial symmetry. In the angular direction, the filters have Gaussian cross-sections, where the standard deviation of the Gaussian controls the filters' directional selectivity. An example of the transfer function for an even-symmetric filter at a single scale and orientation is shown in Figure 5.6(b).

The ratio between the standard deviation and the angular spacing of the filters is a fixed constant, which is set to achieve an even coverage of the 2-D spectrum. This ensures a fixed length-to-width ratio of the filters in the spatial domain. Thus, the cross-section of the transfer function in the angular direction is

$$G(\theta) = \exp\left[-\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2}\right], \tag{5.3}$$

where θ_0 is the orientation angle of the filter, and σ_θ is the standard deviation of the Gaussian spreading function in the angular direction.

Kovesi [86] reported that a filter orientation spacing of 30° provides a good compromise between the need to achieve an even spectral coverage while minimising the number of orientations. Experiments were performed to test the use of more filter orientations, but the additional orientations did not change the quality of the results significantly. For this reason an orientation spacing of 30° is retained for the work reported here.

The outputs from a quadrature pair of filters can be thought of as representing a vector in the complex plane, with the even-symmetric filter output representing the real component, and the odd-symmetric output representing the imaginary component. Convolving an image with even and odd log Gabor filters over a range of scales and orientations therefore produces an array of response vectors, $R_{i,s,\theta}$, that are a function of the pixel position i, scale s and orientation θ. If we let f denote the greyscale image and $W^e_{i,s,\theta}$ and $W^o_{i,s,\theta}$ denote the even-symmetric and odd-symmetric wavelets respectively, the responses of each quadrature pair of filters form the response vector

$$R_{i,s,\theta} = [\Re(R_{i,s,\theta}), \Im(R_{i,s,\theta})] = [f_i \odot W^e_{i,s,\theta}, f_i \odot W^o_{i,s,\theta}]. \tag{5.4}$$

The amplitude of the transform at a given wavelet scale and orientation is given by

$$A_{i,s,\theta} = \sqrt{\Re(R_{i,s,\theta})^2 + \Im(R_{i,s,\theta})^2}, \tag{5.5}$$

and the phase is given by

$$\phi_{i,s,\theta} = \text{atan2}(\Im(R_{i,s,\theta}), \Re(R_{i,s,\theta})). \tag{5.6}$$

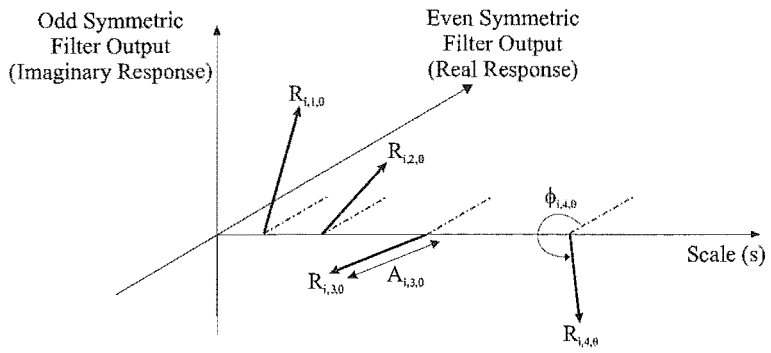


Figure 5.7 Response vectors $R_{i,s,\theta}$ from the convolution of the even and odd log Gabor filters for an arbitrary point in an image at one orientation and multiple scales.

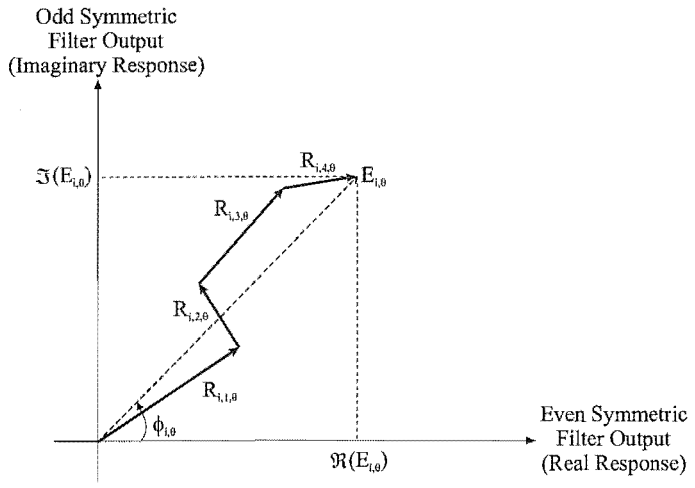


Figure 5.8 Polar diagram showing how the response vectors at a location in the image for four scales and one particular orientation are combined to form $E_{i,\theta}$ (see Equation 5.7).

The response vectors form the basis of the localised representation of the image. Figure 5.7 shows an example of the response vectors at an arbitrary location in an image for one particular orientation and a range of scales.

The response vectors across the range of scales are combined to form

$$E_{i,\theta} = \sqrt{\left(\sum_s \Re(R_{i,s,\theta})\right)^2 + \left(\sum_s \Im(R_{i,s,\theta})\right)^2}. \tag{5.7}$$

Figure 5.8 shows an example of the resultant vector $E_{i,\theta}$ obtained when the response vectors are summed over all scales at a location in the image for one particular orientation. This ensures there is no bias towards features of a particular size which seems appropriate due to the varying size of CLS. The resultant phase $\phi_{i,\theta}$ from the combination of the response vectors is used to determine the feature types within the image. This is described in more detail in the next subsection.

Selection of the filter scales is important in ensuring correct classification of feature types. A wavelength parameter is associated with each scale representing the spatial wavelength of the sine/cosine waveform used in the log Gabor function. To ensure detection of a line-like feature with finite width the wavelength should be set to approximately twice the line width. This makes the parameter dependent on the resolution of the mammogram image. Detection of line-like features of varying size require filters with varying wavelengths (or scales), hence the multiresolution filter bank approach.

5.4.2 Shape parameter calculation

Using the magnitude and phase information from $E_{i,\theta}$ a shape parameter is designed to emphasise line-like structures (CLS) while simultaneously suppressing blob-like features (microcalcifications). The approach allows adaptability to features of varying size and orientation.

Convolution of an image with the log Gabor filter bank (described in the previous subsection) provides a rich set of local property measures whose values differ for various image features. Figure 5.9 gives an illustration of how $E_{i,\theta}$ varies for three different feature types. It should be noted in these examples that the wavelength of the filters in the spatial domain have been chosen large enough to cover the width of each feature. Six orientations have been used in these examples corresponding to a filter orientation spacing of 30°. The centre pixel of each feature on the left hand side of Figure 5.9 provides the $E_{i,\theta}$ vectors on the right hand side. For the blob-like structure, as the orientations change, the magnitude of $E_{i,\theta}$ remains high and the phase remains at approximately 0°. However, for the line-like feature the magnitude of $E_{i,\theta}$ varies considerably, with the phase remaining at 0°. Finally for a pixel on the edge of a step feature $E_{i,\theta}$ varies significantly in magnitude and the phase remains at approximately 90°. From these plots it is clear that all attributes of $E_{i,\theta}$ (magnitude, phase and the dependence of both on orientation) are required to distinguish line-like features from other distinctive feature types.

Observation of how the complex quantity $E_{i,\theta}$ varied throughout an image lead to the definition of a derived real quantity

$$U_{i,\theta} = \Re(E_{i,\theta}) - |\Im(E_{i,\theta})|. \tag{5.8}$$

$U_{i,\theta}$ has a strong positive response to a feature with a positive ridge profile in the orientation being considered. Thus if location i lies on a positive ridge feature and θ is oriented perpendicular to the longitudinal axis of the ridge, $U_{i,\theta}$ is large and positive. Furthermore, $U_{i,\theta}$ decreases strongly as the location is moved perpendicularly away from the axis of the ridge.

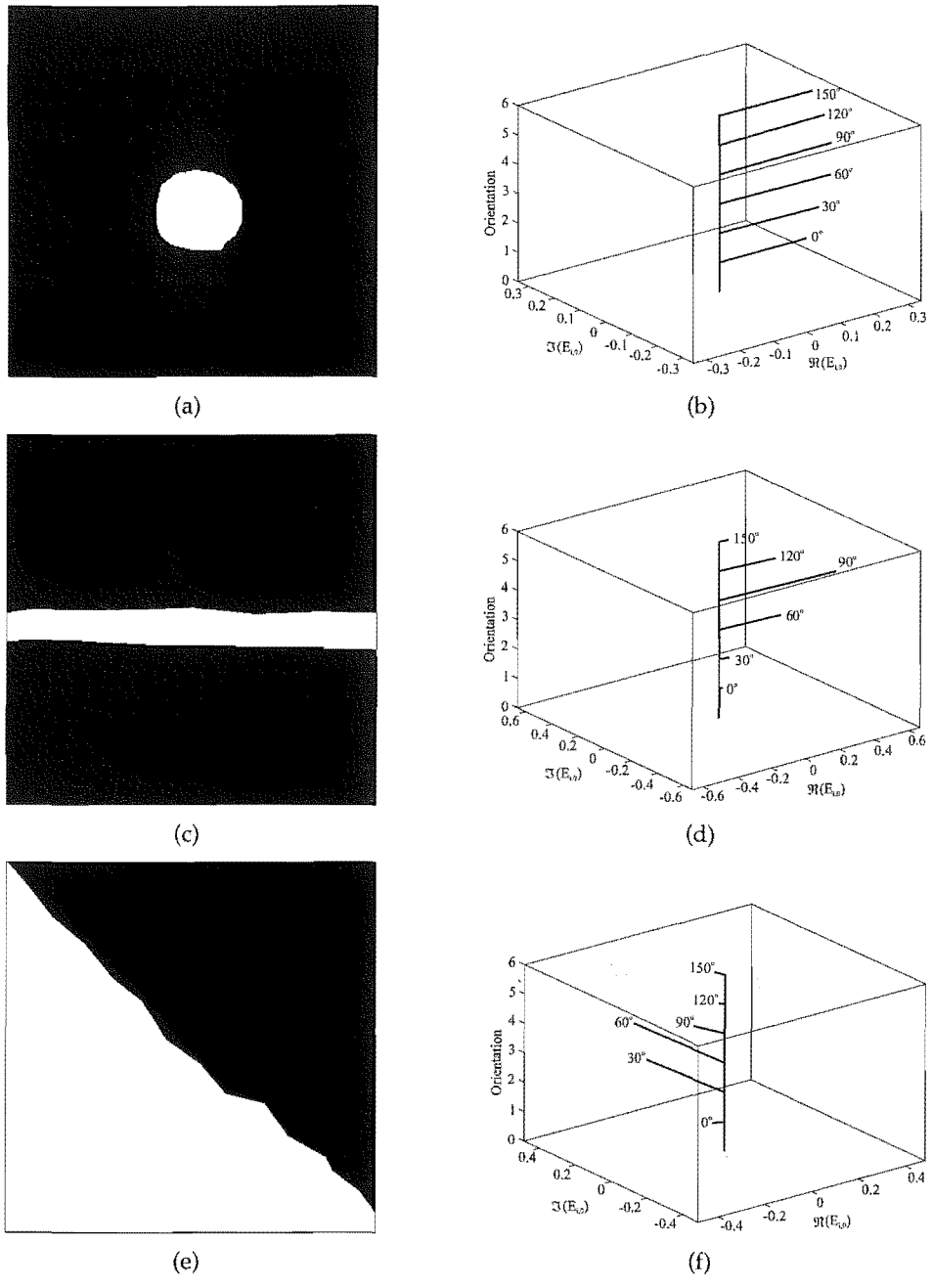


Figure 5.9 Illustration of how $E_{i,\theta}$ varies for different feature types. (a) Binary blob-like feature. (b) $E_{i,\theta}$ values over all orientations for the centre pixels of the blob-like feature. (c) Binary line feature, (c) $E_{i,\theta}$ values over all orientations for the centre pixel of the line-like feature. (d) Binary step edge. (e) $E_{i,\theta}$ value over all orientations for a pixel on the edge of the step feature.

The following two examples are designed to illustrate how $U_{i,\theta}$ highlights positive ridge structures but does not respond strongly to other feature types. Figure 5.10 provides an illustrative example of $U_{i,\theta}$ operating in 1-D using a synthetic signal containing a single positive ridge feature. For convenience the parameter $U_{i,\theta}$ is simply referred to as U in the description of this 1-D example. First, the original signal is convolved with both an even

and odd symmetric log Gabor filter. Figure 5.10 shows a single scale filter with a bandwidth of 1 octave and a spatial wavelength of 34 pixels. This wavelength is set large enough to cover the entire ridge ensuring the signal is not interpreted as two step features. Calculation of parameter U from the convolved signals produces a plot in the spatial domain with a strong positive pulse at the location of the original ridge. Subtraction of the absolute imaginary component of $E_{i,\theta}$ when calculating U ensures a more localised response to the ridge feature. A ringing effect can be seen in this 1-D plot which is a result of the filter response at the start and end of the ridge. As mentioned previously the significance of a feature is only relevant if congruency occurs over a range of frequencies. This is verified by the final plot in Figure 5.10, where U has been summed over 5 scales with a filter spacing that ensures even coverage of the frequency spectrum. The plot shows a smooth response due to cancellation of the ringing effects when summed over multiple scales. At the same time the positive peak in the value of U is reinforced at the location of the ridge due to the consistent response over all filter scales at this location. This example shows that the parameter $U_{i,\theta}$ indeed highlights positive ridge features which are representative of the cross-section of both a line-like and blob-like feature.

To illustrate how parameter U responds to other common feature types Figure 5.11 provides an example of the resulting U when applied to two step features. The step features are created by synthesising a wide pulse representing both a positive step (left hand side) and a negative step (right hand side). A single scale example filter is used in Figure 5.11 with a bandwidth of 1 octave and a spatial wavelength of 10 pixels. This wavelength is chosen small enough to ensure the pulse is not detected as a single feature. Calculation of U from both the even and odd convolved signals produces a signal that is difficult to interpret. It should be noted that in this plot U reaches a minimum value at the location of both step features. After summing over 5 scales (all of which are too small to detect the pulse as a single feature) the resultant U parameter produces two very large negative pulses corresponding to the step features in the original signal. There are only very small positive peaks. This result provides evidence that U does not highlight step edge features.

To reiterate, the purpose of determining a shape parameter is to distinguish line-like features (CLS) from blob-like structures (microcalcifications). This can be achieved by utilising both the results from Equation 5.8 and the orientation information available. The value of the shape parameter is intended to represent the likelihood of a line-like structure at each pixel location i.

Figure 5.10 shows that $U_{i,\theta}$ emphasises signals that are indicative of a positive ridge. This type of signal can be found along the cross section of both a line-like and blob-like structure. It is the variation of $U_{i,\theta}$ over all orientations that provides the important distinction between blob and line features. For a blob-like structure the profile in all orientations is that

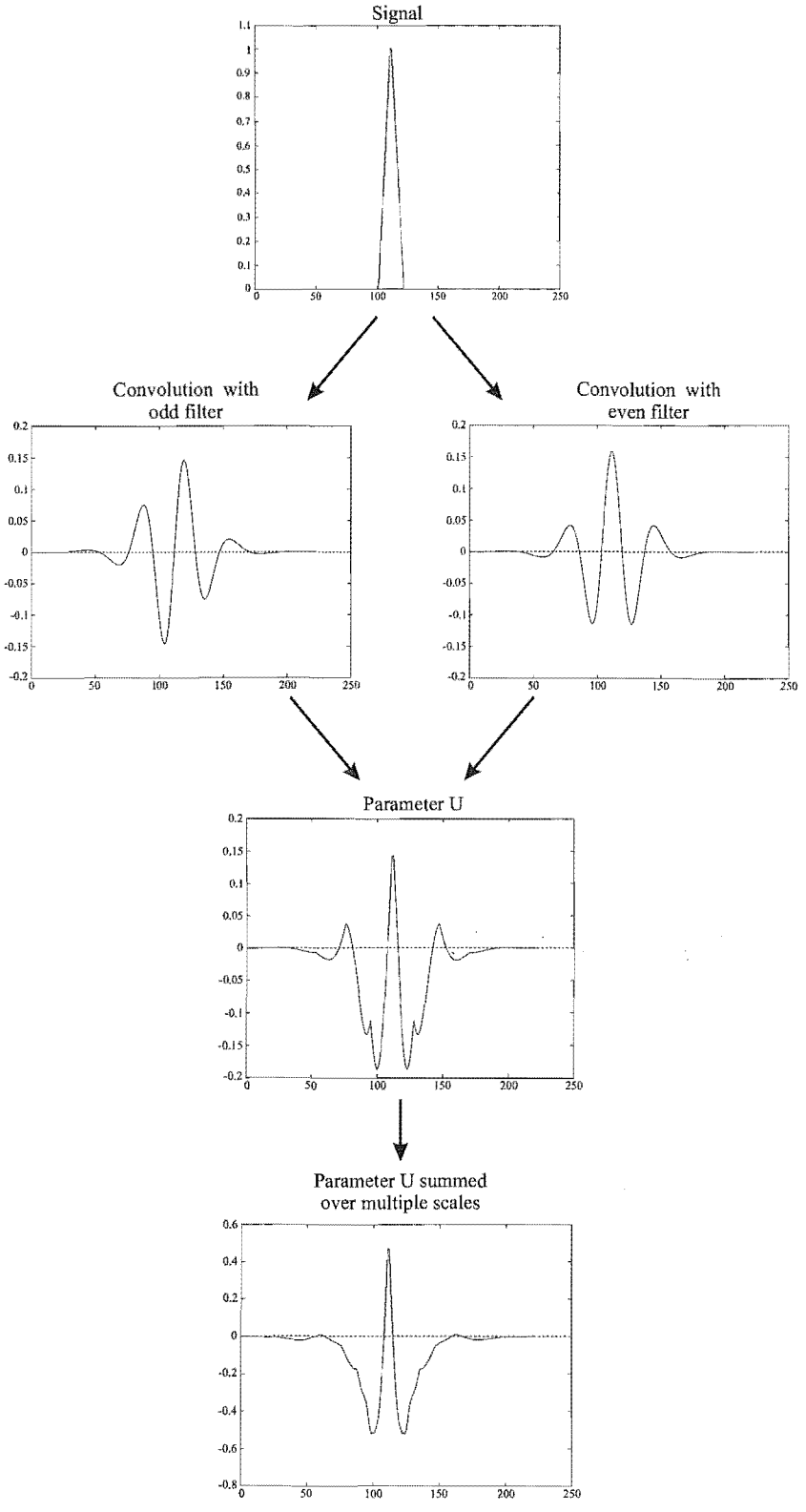


Figure 5.10 Calculation of parameter U in one dimension for a ridge function via convolution with two filters in quadrature.

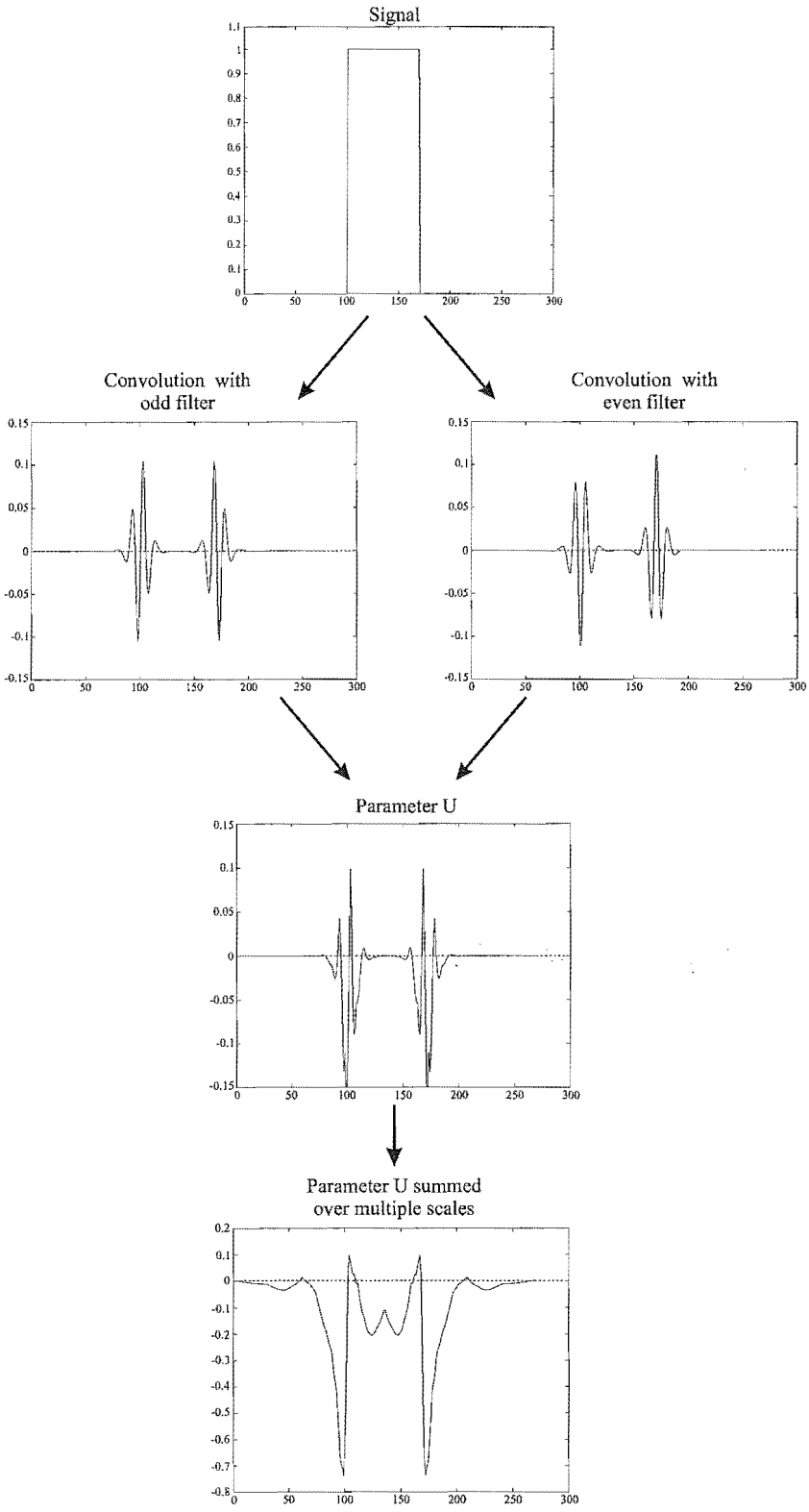


Figure 5.11 Calculation of parameter U in one dimension for two step features via convolution with two filters in quadrature.

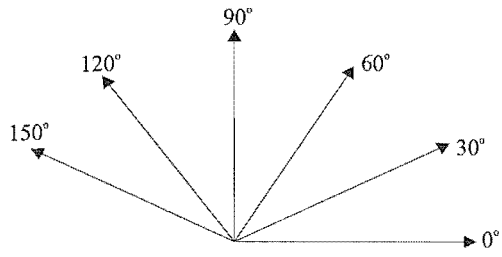


Figure 5.12 Division of the frequency domain when the number of orientations is 6.

of a positive ridge, therefore the value of $U_{i,\theta}$ is high over all orientations. However for a line-like feature only one orientation provides the profile of a positive ridge and $U_{i,\theta}$ is high in only this orientation. This suggests three measures that can be used to help characterise line-like features:

$$X_i = \underset{\theta}{\sigma}(U_{i,\theta}), \qquad M_i = \max_{\theta}(U_{i,\theta}), \qquad OM_i = \underset{\theta}{\text{argmax}}(U_{i,\theta}). \tag{5.9}$$

The parameter X_i is the standard deviation of $U_{i,\theta}$ over all orientations, while parameter M_i is the maximum value of $U_{i,\theta}$ over all orientations. Due to the different properties of the blob and line features mentioned above, line-like features have high values for both X_i and M_i, while blob-like structures have high M_i values but very low X_i values. The measure OM_i is the orientation at which $U_{i,\theta}$ is a maximum. For a line-like feature this orientation is perpendicular to the major axis of the line structure. The importance of determining this orientation is described to follow.

Figure 5.12 illustrates how the orientations of the log Gabor filters are distributed over the frequency domain when the number of orientations is equal to 6. The difference in $U_{i,\theta}$ between blob and line like structures is most dominant at the orientation corresponding to the major axis of each feature (i.e perpendicular to OM_i). At this orientation the profile of a blob-like feature is a positive ridge and has a high $U_{i,\theta}$ value, however along the major axis of a line-like structure the profile is fairly uniform and the value of $U_{i,\theta}$ is low. Let the value of $U_{i,\theta}$ perpendicular to OM_i be denoted as P_i. For example if the maximum $U_{i,\theta}$ value occurs at $OM_i = 30°$, then P_i is the value of $U_{i,\theta}$ at orientation $120°$.

Only the centres of blob-like structures have parameter values which differ significantly from those of line-like features. Therefore it is important to apply the shape parameter only to a limited number of pixels near the centres of the blob-like features. Non-maximal suppression [20] is used to obtain this limited set of pixels. This algorithm is generally only defined in terms of one-dimensional signals and requires an orientation image to be used in conjunction with a feature strength image to determine local maxima. The orien-

tation indicates the direction to scan across the image feature and extract the appropriate one-dimensional signal. This signal is then tested to determine whether the point of interest is a local maximum. In this implementation M_i is non-maximally suppressed in the direction OM_i. M_i is used because it highlights the centre of both line-like and blob-like structures, and OM_i is the orientation of the feature normal [86]. Non-maximal suppression produces a binary image B_i highlighting points that are locally maximum. Using OM_i for the orientation image produces a somewhat quantised result (quantised by the number of filter orientations). However, it has been shown in [86] that increasing the precision of the orientation image does not change B_i noticeably.

Utilising the properties of the above quantities the new phase congruency shape parameter L_i is calculated as

$$L_i = (M_i - P_i)\ X_i\ B_i. \tag{5.10}$$

Multiplication by B_i ensures L_i is only calculated for non-suppressed pixels. The operations performed in this calculation are designed to highlight line-like features while suppressing blob-like structures [102–104]. A high value of L_i indicates that a line-like structure (CLS) at site i is likely. Examples of how the phase congruency shape parameter works on both a simulated binary image and a digitised film-screen mammogram are provided in the next section.

5.5 Results

The purpose of developing a line shape parameter is to provide a distinction between CLS and microcalcifications. In this section a comparison is made between the phase congruency shape parameter and the Hough shape parameter based on their ability to highlight line-like features while simultaneously suppressing blob-like structures.

Figure 5.13 shows both a simulated binary image and part of a digitised film-screen mammogram from FSD-B, each having a size of 480×480 pixels. Throughout this section these images are used to illustrate the operation of each shape parameter method. The binary image in Figure 5.13(a) contains both line-like and blob-like features of varying positioning, size and orientation. This image was designed to provide a simple representation of various CLS and microcalcification properties. The cropped mammogram image in Figure 5.13(b) exhibits numerous line-like features and a large microcalcification cluster. The pixel size in this mammogram is 100μm.

Karssemeijer's Hough shape parameter technique was applied to the original images in Figure 5.13. The success of the Hough method is highly dependent on the size of the moving window (small neighbourhood). The window must be chosen large enough to encom-

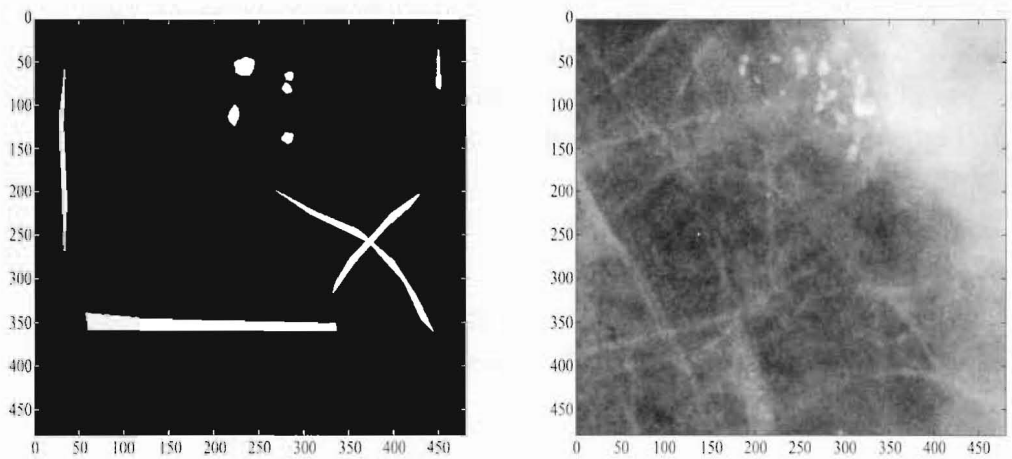


Figure 5.13 (a) Original simulated binary image (b) Original cropped digitised film-screen mammogram image containing a microcalcification cluster.

pass each individual feature while at the same time be small enough to ensure multiple features are not covered simultaneously. After looking at the mammogram images in FSD-B a window size of 15×15 pixels (for a 100μm pixel size image) was chosen to satisfy the above requirements. Figure 5.14 shows the results obtained by applying the Hough shape parameter to the images in Figure 5.13. For the purposes of display the shape parameter values, ζ_i, have been scaled to arbitrary integer units and inverted using [71]

$$\zeta_i' = 255(1 - \frac{\zeta_i}{4}). \tag{5.11}$$

High values of ζ_i' indicate that a line-like structure at i is likely.

From the illustrative examples in Figure 5.14 it is clear that this technique not only highlights line-like structures but also near circular objects in both the simulated and mammogram images. Since the purpose of the shape parameter is to differentiate between these two feature types it appears to be of limited usefulness. Another problem originates from the fixed Hough window size, which causes small features to be blurred dramatically, most obviously in the simulated image. The Hough shape parameter does not represent varying feature sizes well, making it inappropriate for CLS detection.

Also applied to the original images in Figure 5.13 was the phase congruency shape parameter. In the results presented here local frequency information was obtained using 1 octave bandwidth filters over 5 scales and 6 orientations. The wavelength of the smallest scale filter was set to 10 pixels and scaling between successive filters was 1.5. Thus, over the 5 scales, the filter wavelengths were 10, 15, 22, 34 and 50 pixels. The filters were constructed directly in the frequency domain as polar-separable functions: a logarithmic

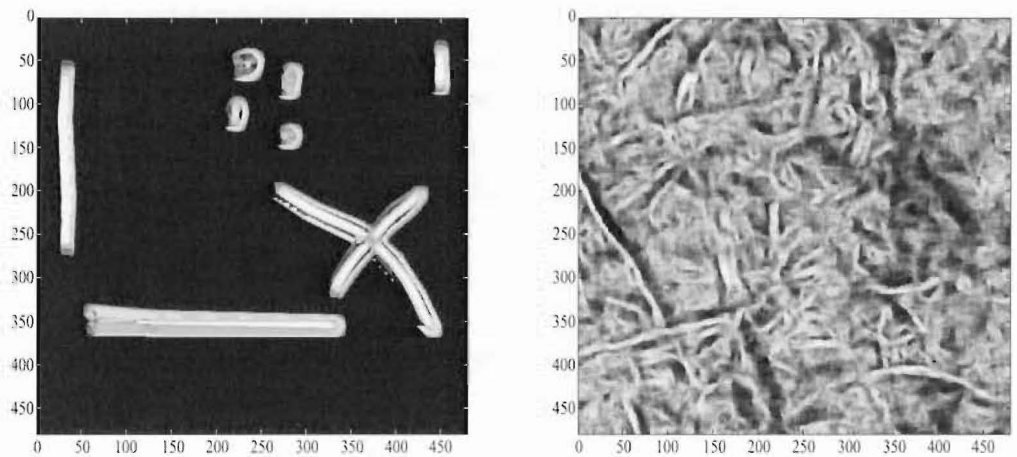


Figure 5.14 Shape parameter image obtained by applying Karssemeijer's Hough technique to (a) the simulated binary image in Figure 5.13(a), and (b) the digitised film-screen mammogram images in Figure 5.13(b).

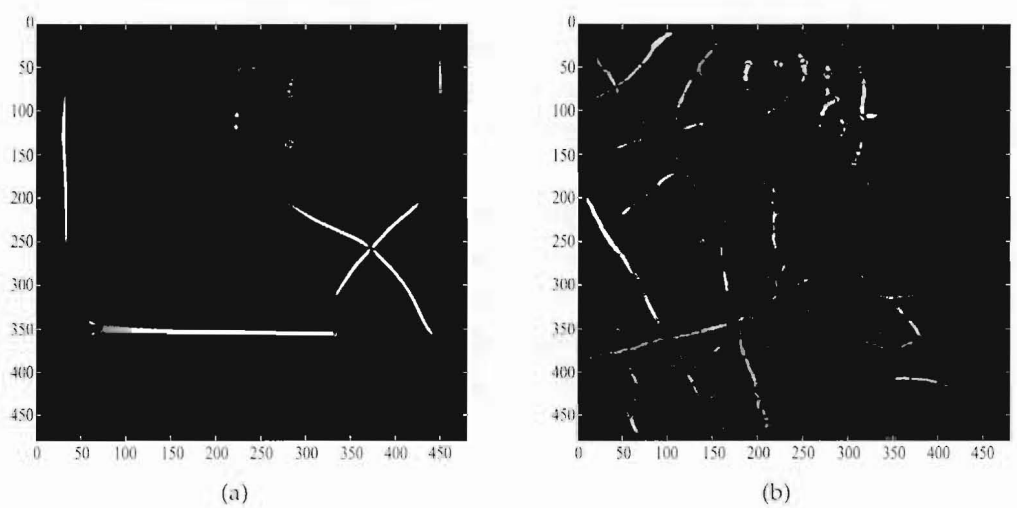


Figure 5.15 Shape parameter image obtained by applying the phase congruency shape parameter to (a) the simulated binary image in Figure 5.13(a), and (b) the digitised film-screen mammogram image in Figure 5.13(b).

Gaussian function in the radial direction and a Gaussian in the angular direction. In the angular direction, the ratio between the angular spacing of the filters and the angular standard deviation of the Gaussians was 1.2. This results in a coverage of the spectrum that varies by less than 1% for the band of frequencies covered by the filters. Note that none of these parameter values are particularly critical. They can be varied by approximately $\pm 10\%$ without considerably affecting the results. Figure 5.15 shows the results obtained by applying the phase congruency shape parameter to the images in Figure 5.13.

In Figure 5.15(a) it is obvious that line-like features are highlighted to a greater extent than blob-like features, indicating an improvement in performance when compared to the Hough shape parameter. The adaptability of the phase congruency shape parameter to varying feature sizes is also demonstrated, giving evidence of the advantage of a multi-scale approach. One problem with the phase congruency method, or at least this implementation of it, is that pixels near the edge of blob-like features have similar properties (with respect to L_i) to those of line-like features making them hard to distinguish.

It is worth mentioning some of the inherent properties of both microcalcifications and CLS that are difficult to interpret for any line-like shape parameter technique. Within a microcalcification cluster individual calcifications are sometimes positioned very close to one another and may even overlap in a certain mammographic view, causing a shape parameter to misinterpret the pixels as belonging to small CLS. The same misinterpretation can occur with slightly elongated microcalcifications which commonly occur in mammography. In both Figures 5.14(b) and 5.15(b) this misinterpretation has caused part of the microcalcification cluster to be highlighted by each shape parameter method. Another cause for inaccurate shape parameter classification results from overlapping CLS. The response at the location of a junction point is often incorrect which can be seen clearly in both Figures 5.14(a) and 5.15(a). It is unclear how to overcome these inherent problems and more work must be carried out to improve performance.

A major difference between the Hough and phase congruency shape parameter techniques is their execution time which was measured using the built-in MATLAB profiler [101]. Implementation of the algorithms in MATLAB is relatively inefficient and neither shape parameter method was optimised for computational efficiency, however the difference in execution time between the two methods is so marked that it seems valid to report the findings. For each mammogram image in FSD-B, a 480×480 pixels subimage was selected. Both shape parameter methods were applied to the data set. The Hough shape parameter had a maximum execution time of 738 seconds, a minimum execution time of 648 seconds and an average of 679 seconds. The phase congruency technique had a maximum execution time of 203 seconds, a minimum of 114 seconds, and an average of 172 seconds. These values show that the phase congruency method executes on average 3.95 times faster than the Hough shape parameter technique. This is of great importance in a computer aided detection system because many cases must be capable of being processed in quick succession.

The phase congruency shape parameter described in this chapter was developed using digitised film-screen mammogram images. In these images CLS appear as high intensity structures against the varying background of the mammogram image. For this reason the shape parameter was designed to detect high intensity line-like features. A simple

modification of the phase congruency shape parameter can be performed for application to raw direct digital mammograms, where CLS appear as low intensity structures against the varying background of the mammogram. In this case the shape parameter would be designed to detect low intensity line-like features.

The reason for developing a line shape parameter is to distinguish CLS from microcalcifications and hence assist in the detection of microcalcification clusters. Chapter 8 provides a comparison between the Hough and phase congruency shape parameter methods based on results from a microcalcification CAD scheme. These results show an improvement in microcalcification detection performance when using the phase congruency shape parameter as input to the CAD system. The phase congruency shape parameter provides better distinction between CLS and microcalcifications ensuring easier classification of the two features, thus leading to more robust computer aided detection.

5.6 Summary

In order to improve the detection performance of a potential CAD scheme identification of CLS is important. A distinction between faint microcalcifications and CLS can only be made on the basis of shape. This chapter detailed the development of a shape parameter designed specifically to highlight CLS while simultaneously suppressing microcalcifications.

Many line-like feature detection methods previously developed for the purpose of mammography are designed specifically for the detection of spicules associated with stellate masses making them inappropriate for CLS detection. Section 5.3 outlined a Hough shape parameter, proposed by Karssemeijer [71], specifically designed for the detection of CLS. Identifying the limitations of this technique provides motivation for the development of a new shape parameter. This is provided in Section 5.4 with the development of a shape parameter based on the method of phase congruency. The inherent properties of CLS such as variable size and orientation support the choice for a multiresolution approach. The phase congruency method uses the local property measures from a multiresolution wavelet filter bank and is designed to highlight line-like features (CLS) while suppressing blob-like structures (microcalcifications).

Section 5.5 discussed the advantages and disadvantages of each shape parameter technique using a simulated binary image and a section of a digitised film-screen mammogram. The results indicate that the phase congruency shape parameter is superior because of its adaptability to the varying size and orientation of CLS and lower execution time.

Chapter 6

Computer aided detection scheme for microcalcifications

6.1 Introduction

The application of Computer Aided Detection (CAD) in digital mammography is an important and rapidly expanding field of research. It has been known for some time that radiologists do not identify all breast cancers that are visible on retrospective review of the mammograms [15,144]. Additionally, the interpretive accuracy of radiologists remains subject to the limitations of human perception. There is evidence that radiologists perform better at detecting subtle microcalcifications when provided with cues from a CAD scheme marking suspicious regions [6, 8, 19, 24, 47, 76, 151]. Moreover, the display for a computer detected microcalcification pattern is found to be useful in interpretation of the spatial distribution of microcalcifications. This distribution is an important cue for distinguishing different cluster types [117]. After extensive investigations and development efforts three CAD systems have been given Food and Drug Administration (FDA) approval (http://www.fda.gov/): ImageChecker (R2 Technology), Second Look (CADx Medical Systems) and MammoReader (Intelligent Systems Software).

Although microcalcification CAD schemes have been reported to have a sensitivity of up to 90%, the false-positive rate associated with the current schemes remains unacceptably high. This chapter describes the development of a CAD system designed to overcome some of the challenges associated with microcalcification detection. The system extends a well known method proposed by Karssemeijer [71,72]. In this scheme a stochastic model is developed to enable the classification of individual pixels within a mammogram into separate classes based on Bayesian decision theory. The extension to the scheme includes

the introduction of a term designed to reduce the number of false positive results associated with CLS. Overall, the CAD scheme aims to mark all suspicious microcalcifications but does not provide a classification into benign or malignant types. The reported methods and results in this chapter are for application to digitised film-screen mammogram images. Adaptation of the CAD scheme to the direct digital modality is mentioned briefly in Section 6.11.

6.2 Existing methods

There has been extensive research carried out in the field of mammographic computer aided detection with specific emphasis on the detection of microcalcification clusters [24–26,113–115,117,141,158]. Unfortunately it is very difficult to assess the relative performance of the methods reported in the literature due to the lack of a common database or common objective criteria. The best results reported to date have a sensitivity of $85 - 90\%$ and a false positive detection rate of approximately 1 false positive per image. Large databases of digital mammogram images are becoming available, which should facilitate direct comparisons of techniques in the future [58].

Reported methods for computer aided detection of microcalcifications fall basically into the following categories:

1. global and local thresholding,
2. multiresolution filtering,
3. artificial neural networks, and
4. Bayesian classifiers.

The best known algorithms for microcalcification CAD spanning this range are reviewed here. Note that Karssemeijer's CAD scheme (based on Bayesian classification) is excluded from this review as it is singled out for further analysis in the remainder of this chapter [71,72].

Microcalcifications have higher mass attenuation coefficients than any other structure in the breast. This suggests that straightforward grey-level thresholding could be a potentially useful method for segmentation [118]. Chan *et al* [24–26] investigates a computer-based method for the detection of microcalcifications using both global and local grey-level thresholding. This work was also mentioned in Chapter 3 in the context of noise equalisation. Initially a difference-imaging technique is applied using linear spatial filters, to enhance the signal from microcalcifications while suppressing the background structure of the breast. A global grey-level threshold is used on the filtered image, extracting possible

microcalcification pixels. To reduce the number of false signals caused by image noise a local grey-level threshold is applied to each potential microcalcification site. This threshold is based on the statistics (mean and standard deviation) of the pixel values in a local region of the mammogram image. Although reasonable microcalcification cluster detection performance has been reported (sensitivity of 82% at a false positive rate of 1 per image), the method suffers from a significant drawback. Adaptive thresholds that are computed locally within small image regions are not only affected by the image noise, but also by image structures like lines and edges. Therefore, in regions with a lot of image structure, thresholds are not adjusted optimally to the noise level and detection performance may deteriorate.

Multiresolution filtering has been applied frequently to the detection of microcalcification clusters [60, 89, 92, 152]. Specifically, wavelets have been recognised as an effective method of enhancing microcalcifications. A well known method developed by Strickland and Hahn uses wavelets as a means of applying matched filters to an image [141, 142]. Their technique implements wavelet transforms to construct a bank of multiscale matched filters employed for the detection of microcalcifications of varying size. The subband images created by four-octave wavelet decomposition are thresholded and combined to yield a map of detection pixels. This map is blurred by convolution with a Gaussian and is then used to weight the subband images before computing the inverse wavelet transform. Individual microcalcifications are greatly enhanced in the output image and segmentation is performed using a global threshold. The method achieves a sensitivity of 55% at a false positive rate of 0.7 per image which the authors admit is a significantly poorer result than those reported for other methods. Reasons for this poor performance include the presence of false positive errors associated with CLS and the detection of other isolated false artifacts. These problems could potentially be eliminated by including an additional algorithm to recognise CLS and a measure that rewards detected pixels occurring in local clusters.

A very sophisticated filter designed by Qian *et al* [125,126] uses a tree-structured nonlinear filter coupled with wavelets. The front end of the method is a cascade of centrally weighted median filters that reduce the image noise while trying to maintain the structure of the microcalcifications. A tree structured wavelet transform is applied to the images employing quadrature mirror filters as basic subunits for both multiresolution decomposition and reconstruction processes. Selective reconstruction of the subimages is used to segment the microcalcifications. The measured sensitivity of this approach is 94% with a false positive detection rate of 1.6 microcalcifications per image. A disadvantage of this technique is that it can fail to preserve the individual microcalcification morphology. Microcalcifications greater than 500μm are often segmented into two small microcalcifications. Furthermore, irregular high intensity microcalcifications greater than 1mm are weakly detected, appearing smaller than their actual size.

Netsch and Peitgen [115] proposed a microcalcification detection scheme using a multiscale analysis based on the Laplacian-of-Gaussian filter and a mathematical model describing a microcalcification as a bright spot of certain size and contrast. First, possible locations of microcalcifications are identified as local maxima in the filtered image on a range of scales. For each detected object, the size and local contrast is estimated based on the Laplacian response (denoted as the scale-space signature). A detection is marked as a microcalcification if the estimated contrast is larger than a predefined threshold which depends on the size of the detected object. The basic method is significantly improved by consideration of the statistical variation of the estimated contrast, which is the result of the complex noise characteristic of the mammograms. The method achieves a sensitivity of 84% at a false positive rate of 1 per image. The author identifies that the majority of false positive errors are associated with CLS within the mammogram and states that an additional algorithm which recognises CLS may reduce the number of false positive clusters detected.

Observation of the vast visual variability of potential microcalcifications in digital mammograms has lead to CAD schemes based on artificial neural networks. Yu and Guan [158] proposed a method for the detection of microcalcification clusters using mixed feature-based neural networks. In the first step, potential microcalcification pixels are detected and grouped into individual microcalcification objects by using mixed features consisting of two wavelet features and two grey level statistical features. A multilayer feedforward neural network classifier generates a likelihood map of potential microcalcification pixels by using the mixed features as inputs. To reduce the number of false positive detections, a second step is included using a set of 31 features extracted from the potential individual microcalcification objects. The discriminatory power of these features is analysed using a general regression neural network. The method achieves a sensitivity of 90% at a false positive rate of 1 per image. The drawback of this technique is the large number of features used to discriminate microcalcifications from their background, thus requiring a large number of images for feature selection (training).

6.3 Karssemeijer's CAD method

The CAD scheme developed by Karssemeijer is one of the best known in the field of microcalcification cluster detection [71,72]. It is based on the use of statistical models and the general framework of Bayesian image analysis.

Implementation of the scheme is performed using the system shown in Figure 6.1. Three different features are used to represent the original mammogram data: the output of a line shape parameter and the local contrast at two different spatial resolutions [71]. Each feature is thought to be important in distinguishing microcalcifications from other structures within the breast. The features are combined using a statistical model to label pixels in the image as microcalcifications, curvilinear structures (CLS) and background.

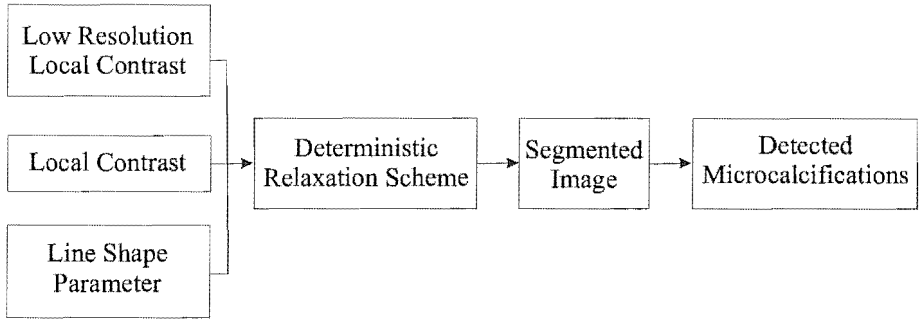


Figure 6.1 Implementation of Karssemeijer's computer aided detection scheme.

The overall detection scheme utilises Bayesian techniques and the application of a Markov random field (MRF) model, in which the latter models that microcalcifications occur in clusters. Starting from an initial segmentation, the labelling of each pixel is optimised by applying an iterative relaxation scheme for updating the pixel labels. Only pixels with a high likelihood of being part of a true microcalcification survive this process. A principal advantage of the approach is that all the information available, i.e. the image data, the current labelling and prior beliefs, is exploited simultaneously.

The output from this CAD scheme is a binary image highlighting the location of individual microcalcifications. As mentioned in Chapter 1 microcalcifications are only considered clinically suspicious when appearing in a cluster of three or more. The spatial distribution of microcalcifications resulting from the CAD scheme are used to determine the clusters present. A detailed description of this clustering process is provided in Chapter 7.

Section 6.4 begins with a description of the three different input features representing the original image data. A statistical model, based on Bayesian methodology and MRF interaction, is detailed in Section 6.5. Section 6.6 describes the iterated updating scheme for estimation of the true segmentation. Section 6.7 introduces an additional MRF term to represent the fact the microcalcifications occur in clusters. An extension to Karssemeijer's scheme is described in Section 6.8 utilising another MRF term designed to reduce the false positives associated with CLS. The determination of all parameter and coefficient settings for the MRF are given in Section 6.9 and example segmentation results are presented in Section 6.10. Finally, adaptation of the CAD scheme to the direct digital modality is mentioned in Section 6.11.

6.4 Feature extraction

To apply a statistical method to the detection of microcalcifications, both the appearance of microcalcifications and the background must be modelled. This is achieved by representing the image data using filtered versions of the original mammogram, which are thought to be important in distinguishing microcalcifications from other structures [71]. Three different features are used to represent the image data: the local contrast at two different spatial resolutions and the output of a line shape parameter.

One of the most distinguishing features between microcalcifications and their background is local contrast. Calculation of the local contrast for both digitised film-screen and direct digital mammograms is described extensively in Chapters 3 and 4 respectively. It is known that the high spatial frequency content and small size of microcalcifications cause the detection accuracy of these objects to depend heavily on the high frequency noise components present in the mammogram. For this reason equalisation of the image noise is applied to the local contrast features resulting in a noise level that is independent of the greyscale. It is the equalised local contrast images that are used as input to the CAD scheme.

A lower resolution local contrast image is included to provide a better representation of the local image intensities. For digitised films it is helpful for identifying scratches and holes in the emulsion, which show up as very sharp features on the mammogram image. By comparing the contrasts at the 2 scales the CAD scheme can better distinguish these artifacts. The low resolution contrast image is simply derived by smoothing the equalised local contrast image using a 3×3 uniform filter kernel (for a 100μm pixel size).

The third feature used to represent the original mammogram is a line shape parameter. This shape parameter is designed to highlight CLS, while simultaneously suppressing microcalcifications and is described extensively in Chapter 5. The use of a line shape parameter is necessary because CLS may easily give rise to false-positive detections in a CAD scheme. Two different shape parameter methods were mentioned in Chapter 5, one based on phase congruency and the other on the Hough transform. Throughout this chapter the phase congruency shape parameter is used.

6.5 Segmentation method

The introduction of stochastic models and Bayesian methods in image analysis has provided a general framework for modelling image data and allowed the inclusion of prior knowledge [13, 35]. In particular Markov random field models have been shown to be appropriate tools for modelling the spatial context of the underlying image, while their use leads to segmentation algorithms which are feasible in practice [34, 36, 93]. These algo-

rithms are basically iterative rules for updating pixel labels in a process of local competition and cooperation. The application of random field models and Bayesian image analysis has been reviewed by Besag [13] and more recently by Dubes and Jain [35].

In this approach to CAD a segmentation X of a mammogram S is estimated by iteratively assigning pixel labels x_i with maximum *a posteriori* (MAP) probability, given the data Y. As mentioned in the previous section the image data at a given site i is represented by three local features: the local contrast at two different spatial resolutions and the output of a line shape parameter. The values of the three features at a particular site i are represented by the vector $\mathbf{y}_i$. Additionally, three pixel classes are distinguished: background, microcalcifications and CLS, with labels respectively numbered as $k = 1, 2$ and 3. (Note that Karssemeijer incorporated a 4th class, image artifacts [71], but the quality of the images used in this study made this unnecessary.)

According to Bayes' theorem, given a prior distribution $p(X)$ and the conditional density $p(Y|X)$ the posterior distribution of X given Y is related by

$$p(X|Y) \propto p(Y|X)p(X). \tag{6.1}$$

In practice the logarithm of the above expression is evaluated because it requires less computation,

$$\ln p(X|Y) \propto \ln p(Y|X) + \ln p(X). \tag{6.2}$$

The probability $p(X)$ incorporates prior knowledge of the spatial interaction between pixels and the conditional density $p(Y|X)$ models how well a certain segmentation explains the given data. As shown in Equation 6.1, Bayes' theorem combines these two terms determining the posterior distribution $p(X|Y)$ which indicates the likelihood of a segmentation X given the data Y. Suitable models for both the density function $p(Y|X)$ and the prior distribution $p(X)$ are adopted in order to maximise the posterior probability of X. A description of both these models follows.

A Markov random field (MRF) is used here as a model for $p(X)$. It is defined in terms of the local dependence of each individual pixel label x_i on the labels x_{δ_i} in its neighbourhood. The probability distribution corresponding to a MRF is given by a Gibbs density which has the form

$$p(X) = \frac{1}{Z} \exp[-U(X)] \tag{6.3}$$

where $U(X)$ is the energy function and Z is a constant of normalisation known as the partition function. Observation of Equation 6.3 indicates that maximising $p(X)$ corresponds to the minimisation of $U(X)$. This property is utilised in Section 6.6.

The notion of neighbourhood is central to MRF models and the relationships within these

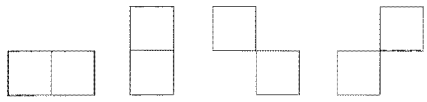


Figure 6.2 Pairwise interaction models for the MRF.

neighbourhoods are imposed in a spatially meaningful way. If δ_i is the neighbourhood of a pixel at site i in a MRF, then the conditional probability satisfies the following

$$p(x_i|x_j, \forall j \neq i) = p(x_i|x_j, j \in \delta_i). \tag{6.4}$$

In this implementation the neighbourhood is restricted to the class of pair-wise interaction terms as shown in Figure 6.2. This means the neighbourhood system δ_i at site i is defined as the eight nearest neighbour pixels surrounding site i.

With this in mind the energy function is given by

$$U(X) = \sum_{i=1}^{N} A(x_i) + \sum_{\langle ij \rangle} B(x_i, x_j) \tag{6.5}$$

in which $\sum_{\langle ij \rangle}$ denotes summation over all neighbourhood pairs of sites in S and N is the total number of sites. A neighbour pair $\langle ij \rangle$ is by definition a pair for which j is an element of δ_i. The interaction coefficient $B(x_i, x_j)$ models the *a priori* likelihood of a particular pairing of labels, and is assumed to be independent of the absolute positions of the sites i and j in S. Also, the external field $A(x_i)$ is taken to be space independent and is used for adjusting the sensitivity of the segmentation algorithm for each class.

From Equations 6.3 and 6.5 the conditional probability of a label to occur at a particular site i is

$$p(x_i = k|x_{\delta_i}) \propto \exp\left[-A(k) - \sum_{n=1}^{K} B(k, n) G_{\delta_i}(n)\right], \quad k = 1, 2, \ldots, K, \tag{6.6}$$

where the number of neighbours labelled as n is denoted by $G_{\delta_i}(n)$ and K is the number of classes, which in this case is equal to three. This model is determined by comparing the energy functions $U(X)$ for segmentations which only differ at site i.

Now consider the conditional density $p(Y|X)$. In modelling the likelihood of the data the following probability density functions are used

$$f(\mathbf{y}_i|x_i) = f(c_i', c2_i', L_i \mid x_i) \tag{6.7}$$

where c_i' is the equalised local contrast value (see Chapters 3 and 4), $c2_i'$ is the low resolution equalised local contrast, and L_i is the line shape parameter indicating whether a line or blob-like feature at site i is likely (see Chapter 5). The values of these three local features at a particular site i are represented by the vector $\mathbf{y}_i$.

Each of the densities associated with the local features is assumed to be normally distributed. This assumption is verified to be true in good approximation in Section 6.9. For each label ($k = 1, 2, 3$) there exists two vectors $\sigma(k)$ and $\mu(k)$ representing the standard deviation and mean respectively of the local feature densities. Assuming the standard deviation $\sigma(k)$ and the mean $\mu(k)$ are known, the conditional density has the form

$$p(\mathbf{y}_i|x_i = k) \propto \exp\left[-\left\|\frac{\mathbf{y}_i - \mu(k)}{2\sigma(k)}\right\|^2\right]. \tag{6.8}$$

6.6 Iterated conditional modes

A common criterion for pixel labelling is the maximum *a posteriori* (MAP) estimate, which is obtained by choosing an estimate for the labelling X that maximises the posterior probability $p(X|Y)$

$$X_{MAP} = \underset{X}{\text{argmax}}(p(X|Y)). \tag{6.9}$$

It is possible, in principle, to compute the right-hand side of Equation 6.9 and find the global optimum. The computation, however, is enormous and thus impractical. For example, given an $M \times M$ pixel image with K labels the posterior probability must be maximised over K^{M^2} possible labellings.

Three well known iterative updating schemes proposed in the literature that aim at MAP estimates are simulated annealing (SA) [49, 82], maximiser of posterior marginals (MPM) [13] and iterated conditional modes (ICM) [13]. A relative comparison of the above three algorithms has indicated that ICM provides good segmentation and is the most robust if a good initial segmentation is available [35]. For this reason the ICM algorithm is the iterative scheme used here to try to find the true segmentation $\hat{X}$.

Besag [13] proposed the ICM method as a computationally feasible alternative to the MAP estimation. The method attempts to estimate pixel labels in a computationally simple manner while retaining the MRF as a model of prior information. Given the local dependence of the MRF, realisations of $p(X)$ tend to consist entirely of a single label. Of course, this is not true of $p(X|Y)$ because of the influence of the data. The ICM algorithm is designed to avoid the tendency of a MRF to degenerate into a single label image.

In this scheme pixel labels are updated by each time choosing the most probable label x_i'

given the current estimate of the rest of the labelling $X_{S\setminus i}$ (excluding site i) and the data Y

$$x_i' = \underset{k}{\operatorname{argmax}}[p(x_i = k|Y, X_{S\setminus i})]. \tag{6.10}$$

Visiting each site at random the iterative process converges to a local maximum of the posterior distribution $p(X|Y)$. Usually only a few iteration cycles are needed for convergence. Using Bayes' relation (Equation 6.1) and the Markov property the probability to be maximised can be written as

$$p(x_i = k|Y, X_{S\setminus i}) \propto p(Y|x_i = k, X_{S\setminus i})p(x_i = k|x_{\delta_i}), \quad k = 1, 2, \ldots, K, \tag{6.11}$$

where x_{δ_i} is the current labelling of the pixels in the neighbourhood δ_i. The labels surrounding site i are considered to be so uncertain that it is better to omit them from the conditioning set of the data term in Equation 6.11. This restricts the influence of the current neighbour labels to the interaction term, which is defined by Equation 6.6. Substituting Equations 6.8 and 6.6 into Equation 6.11 gives

$$p(x_i = k|Y, X_{S\setminus i}) \propto \exp\Big[-\left\|\frac{\mathbf{y}_i - \boldsymbol{\mu}(k)}{2\sigma(k)}\right\|^2 - A(k) - \sum_{n=1}^{K} B(k,n)G_{\delta_i}(n)\Big], \quad k = 1, 2, \ldots, K. \tag{6.12}$$

As mentioned in Section 6.5 it is common practice to evaluate the logarithm of the above expression as it requires less computation. The label x_i corresponding to the maximum value for $p(x_i = k|Y, X_{S\setminus i})$ is unaffected by taking the logarithm. Maximising the conditional probability is equivalent to minimising a new energy function

$$E(x_i = k) = \left\|\frac{\mathbf{y}_i - \boldsymbol{\mu}(k)}{2\sigma(k)}\right\|^2 + A(k) + \sum_{n=1}^{K} B(k,n)G_{\delta_i}(n), \quad k = 1, 2, \ldots, K. \tag{6.13}$$

The ICM method can only be applied when a good initial estimate of the segmentation is available. Such an estimate is obtained by maximum likelihood (ML) which maximises the data term in Equation 6.12 without considering neighbourhood pixel labels.

6.7 Long range interaction

A long range interaction term is introduced to utilise the prior knowledge that microcalcifications occur in clusters. Faint isolated spots should not be interpreted as microcalcifications. The long range model implemented by Karssemeijer [71] is an extension of

Equation 6.6 to give

$$p(x_i = k|x_{\Delta i}) \propto \exp\left[-A(k) - \sum_{n=1}^{K} B(k,n)G_{\delta_i}(n) - \sum_{n=1}^{K} B'(k,n)G_{\Delta i}(n)\right] \quad k = 1, 2, \ldots, K, \tag{6.14}$$

in which $G_{\Delta i}(n)$ is the number of pixels labelled as n in a large neighbourhood of i, denoted Δi. The coefficients $B'(k,n)$ are chosen such that only those corresponding to microcalcification pairs are non-zero and therefore contribute to the interaction sum. This simplifies Equation 6.14 to

$$p(x_i = k|x_{\Delta i}) \propto \exp\left[-A(k) - \sum_{n=1}^{K} B(k,n)G_{\delta_i}(n) - B'(k)N_{c,i}\right], \quad k = 1, 2, \ldots, K, \tag{6.15}$$

with $N_{c,i}$ the number of microcalcification sites in Δi.

In calculating the long range interaction term, all sites within 30 pixels distance are considered, making up a total of about 2800 distant neighbours. This value was chosen based on the average distance between neighbouring microcalcifications expected in practice (i.e. $20 - 25$ pixels for a 100μm resolution image). The neighbourhood is large enough to encompass multiple microcalcifications ensuring the ability to detect the presence of a cluster.

The long range interaction coefficient specified by Karssemeijer [71] is calculated as

$$B'(2) = -\frac{250}{N_{\Delta i}}, \tag{6.16}$$

where $N_{\Delta i}$ is the total number of pixels in the large neighbourhood Δi. In dense microcalcification clusters the long range interaction term highly influences the conditional probability given in Equation 6.15. To prevent an unrestrained increase in the number of pixels labelled as microcalcifications in this case, it is necessary to limit the strength of the field $N_{c,i}$ representing the number of nearby microcalcification pixels. A cut-off at 1% of the maximum value in $N_{\Delta i}$ is applied. With the coefficient specified as in Equation 6.16, this equates to a maximum magnitude for the long range interaction term of 2.5, roughly equal to the values specified for other terms in the MRF (see Section 6.9).

6.8 Connectivity

Sections 6.3 to 6.7 inclusive have detailed the microcalcification CAD scheme proposed by Karssemeijer [71, 72]. This section describes an extension to Karssemeijer's technique, introducing an additional MRF term, referred to as the connectivity term, in conjunction with the line shape parameter in an attempt to reduce the number of false positive results associated with CLS.

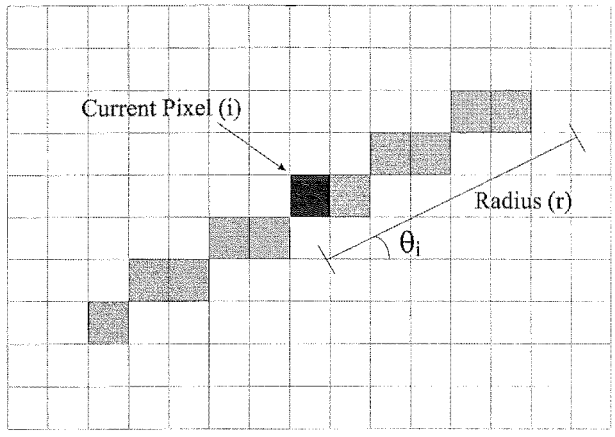


Figure 6.3 Connected line l_i representing the pixels involved in calculation of the connectivity term at pixel i, for a particular radius $r = 5$ and orientation θ_i.

The connectivity term makes use of the knowledge that pixels classified as curvilinear structures are expected to occur as connected lines in the segmented image. In calculating this term, information is required about the orientation θ of potential line-like features. Pixels associated with any line-like structure within the image are expected to have an orientation aligned with the structure's axis. As mentioned in Chapter 5 the phase congruency shape parameter provides a feature orientation at each pixel location with value ranging from $0-180°$. It is this orientation that is used in computing the connectivity term. If a curvilinear structure is present at a given site i, the orientation is along the longitudinal axis of this feature, whereas in other cases the image noise can be expected to generate a random orientation.

Determining the feature orientation at a site i allows the construction of a connected line l_i passing along the major axis of any feature present. The set of pixels associated with l_i are determined using an Octantal digital differential analyser (Octantal DDA) algorithm described in [5]. The length of this connected line is determined by a radius parameter r which sets the distance either side of the current pixel i. The radius value is chosen such that it is greater than any expected microcalcification radius however not so large that it considerably increases the computational load of the CAD scheme. In this implementation r is set to 20 pixels for a 100μm resolution image, corresponding to a distance of approximately $0.8mm$. Figure 6.3 shows an example connected line l_i for a pixel in an image with orientation θ_i and radius $r = 5$. Only the set of pixels associated with l_i influence the connectivity term for pixel i.

In the same manner as the long range interaction term a connectivity term is included as

an extension of Equation 6.15, thus

$$p(x_i = k|x_{\Delta i}) \propto \exp\left[-A(k) - \sum_{n=1}^{K} B(k,n)G_{\delta_i}(n) - B'(k)N_{c,i} - B''(k)N_{l,i}\right], \quad k = 1,2,\ldots,K, \tag{6.17}$$

where $N_{l,i}$ is the number of curvilinear structure sites in the connected line l_i. The coefficients $B''(k)$ are chosen such that only that for the curvilinear structure class is non-zero and contributes to the conditional probability.

The coefficient for the connectivity is set to

$$B''(3) = -\frac{1}{M}, \tag{6.18}$$

where M is the total number of pixels in the connected line l_i. Thus if all pixels in l_i are labelled as belonging to a curvilinear structure ($k = 3$) the value of $B''(3)N_{l,i}$ is equal to -1, indicating a high likelihood of a curvilinear structure.

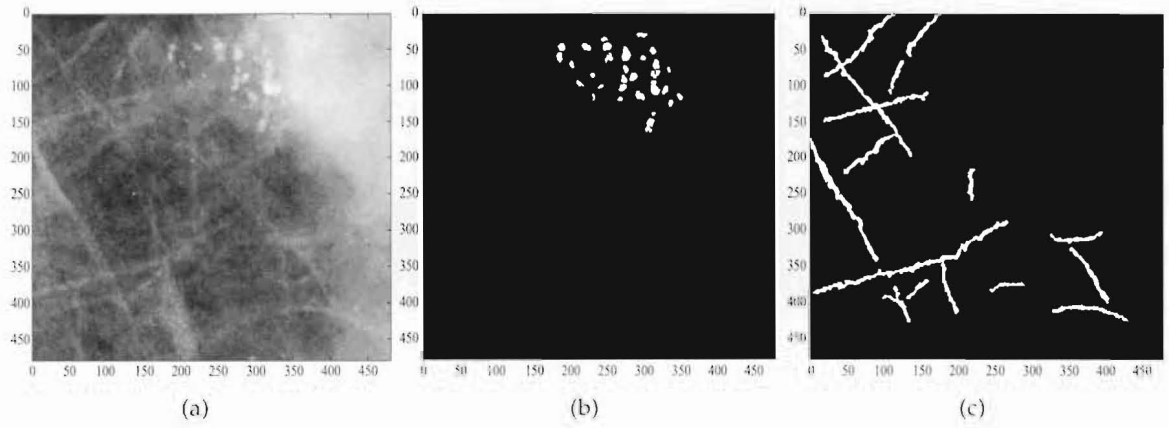
Various other methods of applying connectivity were investigated including the use of a weighting dependent on the orientation of all pixels in the connectivity line l_i. The weighting involved a summation of the difference in orientation between each pixel in l_i and the current pixel i. The smaller the summation the higher the likelihood of the current pixel being associated with a curvilinear structure. It was found however that inclusion of this additional weighting term did not improve the final CAD result.

With the inclusion of both the long range interaction term and the connectivity term the new energy function to be minimised by the ICM algorithm is

$$E(x_i = k) = \left\| \frac{\mathbf{y}_i - \boldsymbol{\mu}(k)}{2\sigma(k)} \right\|^2 + A(k) + \sum_{n=1}^{K} B(k,n)G_{\delta_i}(n) + B'(k)N_{c,i} + B''(k)N_{l,i}, \quad k = 1,2,\ldots,K. \tag{6.19}$$

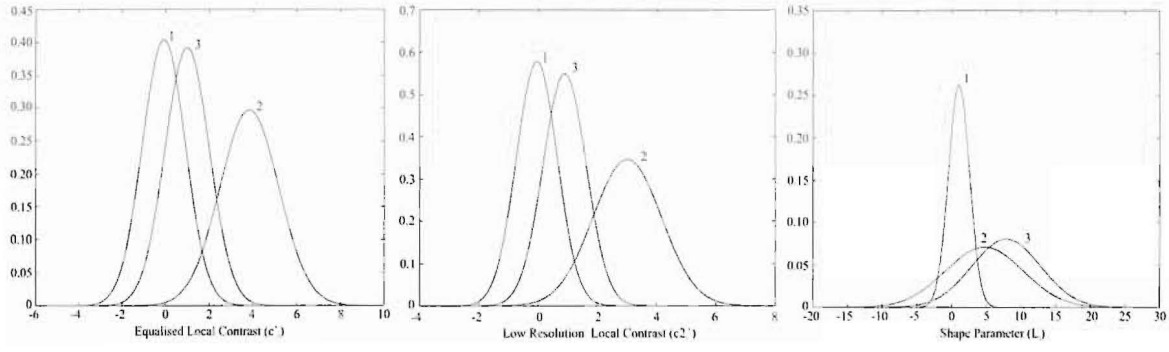
6.9 Random field parameters and probability densities

To obtain the results presented in the next section the parameters of the conditional probability density functions $f(c1'_i, c2'_i, L_i \mid x_i)$ were estimated from 4 mammograms in FSD-B in the following way. A subimage (480 × 480 pixels in size) of each of the 4 mammograms was selected such that it contained microcalcifications, curvilinear structures and background. The four subimages were manually segmented to determine the pixels belonging to each of the three classes. Figure 6.4 shows an example of an original mammogram subimage and its corresponding manually segmented microcalcification pixels and curvilinear structure pixels. All other pixels in the image are assumed to be background.



(a) (b) (c)

Figure 6.4 (a) Original mammogram subimage (b) Manually segmented microcalcification pixels. (c) Manually segmented curvilinear structure pixels.



(a) (b) (c)

Figure 6.5 Estimated conditional densities corresponding to (a) local contrast $f(c_i'|x_i)$, (b) low resolution local contrast $f(c2_i'|x_i)$ and (c) line shape parameter $f(L_i|x_i)$. The labelling $k = 1$ corresponds to the background pixels, $k = 2$ to the microcalcification pixels and $k = 3$ to the curvilinear structures.

From the 4 manually segmented mammogram subimages the conditional probability density functions for each class were estimated for all local features. Modelling the conditional probability densities with normal distributions appeared to be valid upon visual inspection. The fit to a normal distribution was best when the number of pixels associated with a particular class was high. Representative values for the mean and standard deviation of each normal distribution were obtained by averaging the corresponding values calculated over the 4 segmented images. Figure 6.5 shows the estimated probability density functions where the labelling ($k = 1, 2, 3$) represents the background, microcalcification and curvilinear structure distributions respectively as before.

Table 6.1 Parameters of the conditionally independent normally distributed probability densities $f(\mathbf{y}_i|k)$, and coefficients of the random field model. Labels correspond to background $k = 1$, microcalcification $k = 2$, and CLS $k = 3$.

	c_i'		$c2_i'$		L_i		MRF model			
k	μ	σ	μ	σ	μ	σ	A(k)	B(k,1)	B(k,2)	B(k,3)
1	-0.08	0.99	-0.05	0.69	1.04	1.51	0	0	0.2	0.5
2	3.84	1.34	3.00	1.15	4.73	5.59	4	0.2	0	2
3	0.99	1.02	0.89	0.72	7.88	4.95	2	0.5	2	0

It appears that the probability densities of local contrast values have little overlap for background and microcalcifications, which suggest that the detection of microcalcifications would be fairly simple. However it should be noted that the number of background pixels is overwhelmingly large compared to the number of microcalcification pixels and using a simple threshold for both c_i' and $c2_i'$ would lead to a large number of false positive pixels per image.

Table 6.1 shows the estimated parameters from the conditional density distributions in Figure 6.5. It was found that slight changes in these parameter values ($\pm 10\%$) did not considerably affect the CAD results, provided the general trends in the distributions were upheld. For example, the mean distribution value for the line shape parameter should be highest for pixels corresponding to CLS, followed by microcalcifications and then background. Also shown in Table 6.1 are the random field coefficients $A(k)$ and $B(k,n)$. Suitable values for these random field coefficients were experimentally determined from the mammogram images in FSD-B as follows.

The external field coefficients $A(k)$ were chosen such that pixels have a high likelihood of being labelled as background, followed by curvilinear structures then microcalcifications. This is to be expected from the frequency of occurrence of each of these classes in a mammogram image.

For the pairwise interaction coefficients $B(k,n)$, spatial continuity of the classes to be labelled is assumed. The value of the coefficient $B(2,3)$ modelling the interaction between microcalcifications and curvilinear structure labels is chosen high to disallow patches consisting of a mixture of these two classes. Interaction with the background class is chosen small to allow microcalcifications consisting of only one or a few pixels to survive if they are bright enough. It should be noted that the CAD scheme is not extremely sensitive to changes in both $A(k)$ and $B(k,n)$. These values can vary by approximately $\pm 15\%$ without considerably affecting the results.

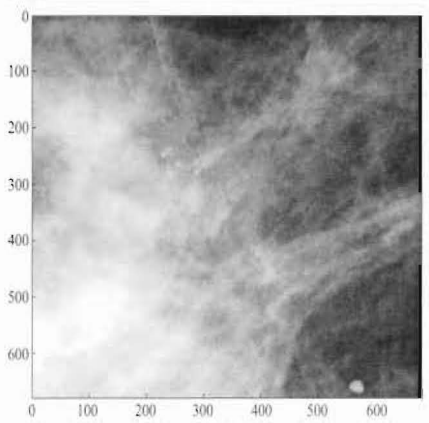


Figure 6.6 Example mammogram subimage from FSD-B.

6.10 Results

The examples presented in this section illustrate the operation of the CAD scheme when applied to selected mammogram images in FSD-B. A fixed number of 7 iteration cycles has been used, leading to convergence of the microcalcification pattern in most cases, or at least convergence to within 0.01%.

Figure 6.6 shows an example mammogram subimage (680×680 pixels in size) from FSD-B, which is used to illustrate the operation of the CAD scheme to follow. This subimage provides clear visualisation of both curvilinear structures and a single microcalcification cluster.

Figure 6.7 shows an example of the CAD results obtained for the mammogram image in Figure 6.6. The binary images in Figures 6.7(a) and (b) correspond to the microcalcification segmented objects after 1 iteration cycle and 7 iteration cycles respectively. Also shown in Figures 6.7(c) and (d) are the curvilinear structure segmented objects after 1 and 7 iteration cycles respectively. It is clear from this figure that microcalcifications are adequately distinguished from curvilinear structures and background, and the morphology of the microcalcifications is preserved. The detection of curvilinear structures does not appear to be as convincing, but note that the main objective of the CAD scheme is to detect microcalcifications and not curvilinear structures. The coefficients in the MRF model were chosen to achieve this objective.

By comparing the results in Figures 6.7(b) and 6.8 the effect of the long range interaction can be judged. Figure 6.8 shows a decrease in microcalcification detection sensitivity within the cluster when long range interaction is omitted. Using the interaction term ensures

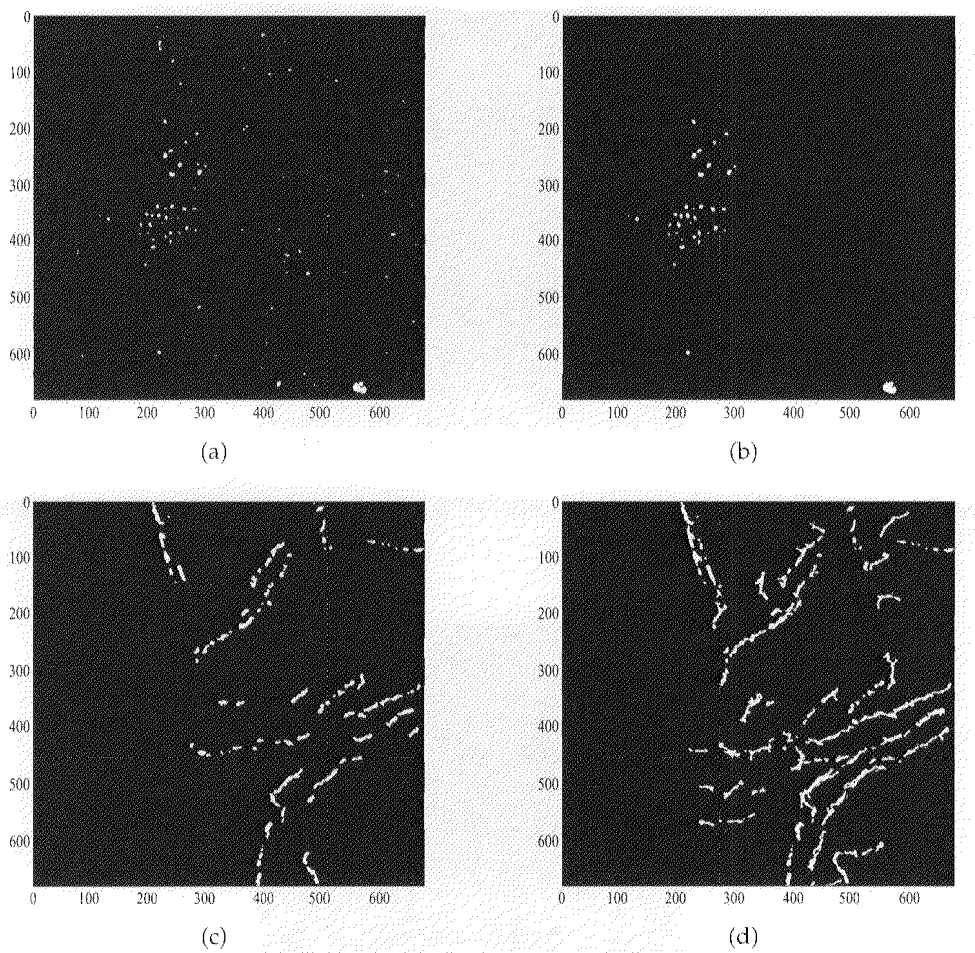


Figure 6.7 CAD results for the mammogram image in Figure 6.6. (a) Microcalcification segmentation after 1 iteration cycle. (b) Microcalcification segmentation after 7 iteration cycles. (c) Curvilinear structure segmentation after 1 iteration cycle. (d) Curvilinear structure segmentation after 7 iteration cycles.

the sensitivity inside the cluster is high enough to detect very subtle microcalcifications as small as one pixel in size, whereas outside the cluster false positive detection is avoided. A few bright isolated spots are detected in Figure 6.8, whereas in Figure 6.7(b) these are not mistaken for microcalcifications because of their isolated positions.

Figure 6.9 shows the microcalcifications detected when the connectivity term is excluded from the MRF. Exclusion of this term does not effect the labelling of the microcalcification cluster however it does produce more false positive detections associated with curvilinear structures within the image. Just to reiterate, this term is included as an extension to Karssemeijer's CAD scheme and offers an improved detection performance by reducing the number of false positive detections related to CLS.

Further examples of the segmented output from the CAD scheme are shown in Figure 6.10.

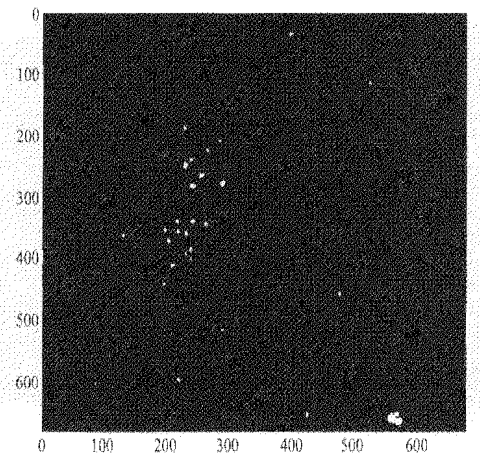


Figure 6.8 Microcalcifications detected after applying the CAD scheme to the image in Figure 6.6(a) without the long range interaction term.

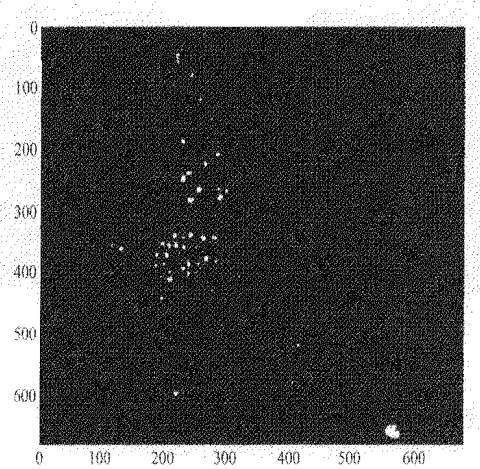


Figure 6.9 Microcalcifications detected after applying the CAD scheme to the image in Figure 6.4(a) without the connectivity parameter.

These subimages are obtained from FSD-B. They illustrate the correct operation of the CAD scheme on a range of mammogram images. A more detailed performance evaluation of the microcalcification detection scheme is provided in Chapter 8.

6.11 Application to direct digital mammogram images

All results and examples in this chapter have been obtained using digitised film-screen mammogram images. For application to direct digital mammograms the only change required to the CAD scheme is in the statistics of the conditional probability density functions for each input local feature. The MRF coefficients including the nearest neighbour terms and the external field coefficients are dependent on the spatial interaction between pixels

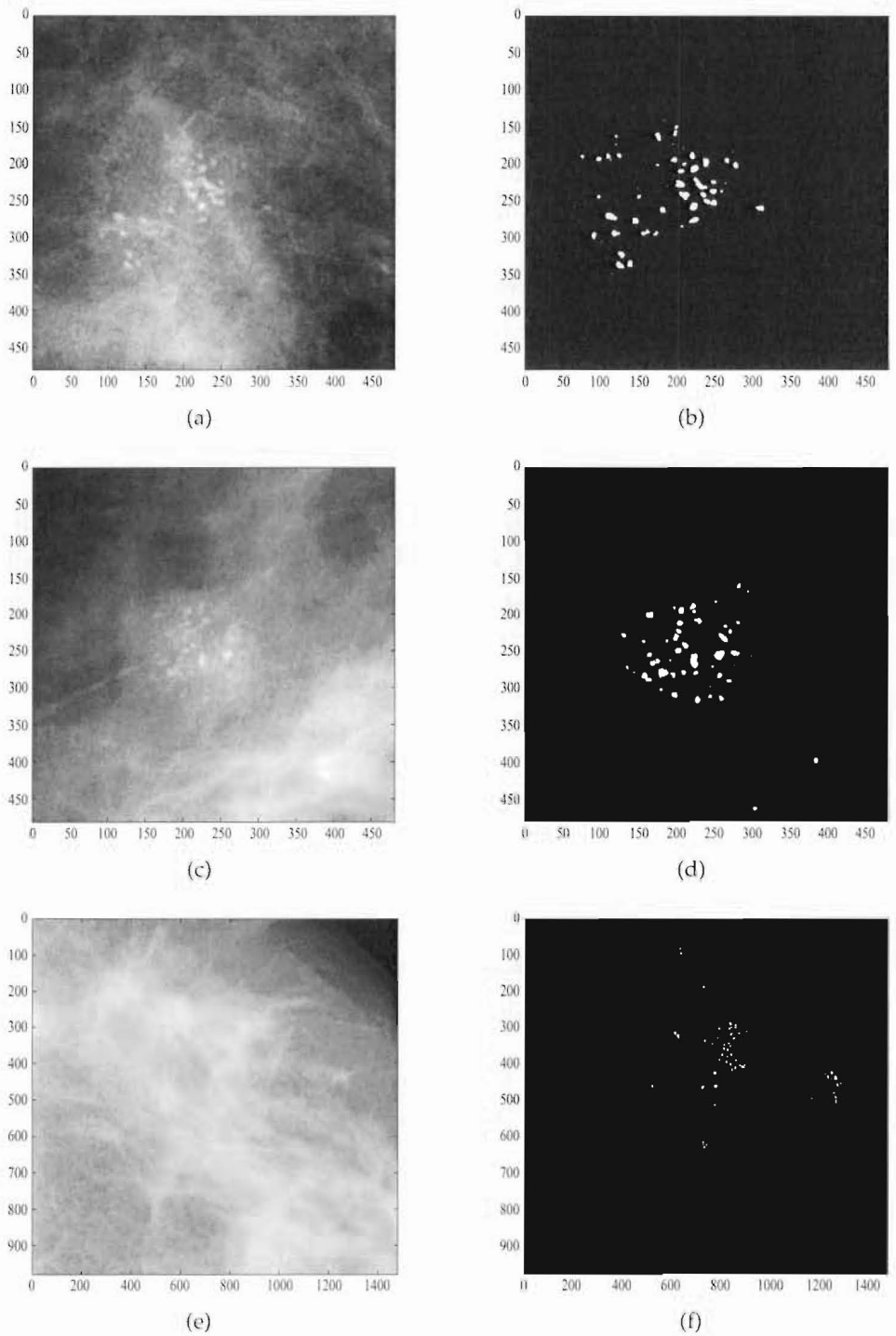


Figure 6.10 Example CAD results. (a, c, e) Original mammogram subimages from FSD-B. (b, d, f) Detected microcalcification pixels.

and not on the modality used. This is also the case for the long range interaction term and the connectivity parameter.

6.12 Summary

This chapter demonstrated the application of a computer aided detection scheme based on Bayesian image analysis for the detection of microcalcification clusters. A simple iterative updating scheme (ICM algorithm) was used to perform this complex recognition task. This makes use of local image features to represent the image data.

To increase the likelihood of microcalcifications occurring in clusters, long range interaction between pixel labels was modelled in the prior distribution $p(X)$, in addition to nearest neighbour interaction stimulating spatial continuity. The long range interaction ensures the sensitivity inside a cluster is high enough to detect very subtle microcalcifications while at the same time removing isolated false microcalcification detections.

The inclusion of a line shape parameter lead to an efficient way of distinguishing curvilinear structures from faint microcalcifications. This distinction is improved with the inclusion of a connectivity term modelled in the prior distribution $p(X)$. The connectivity utilises the fact that pixels classified as CLS are likely to occur along connected lines within the image. This improves microcalcification detection performance by reducing the number of false positive detections related to CLS.

By iterative optimisation of the pixel labels, classification is performed by combining all evidence at hand. In this process even microcalcifications as small as one pixel can survive, provided the evidence for their existence is high enough. With the current quality of mammographic techniques, microcalcifications smaller than 200μm may be perceptible. These should not be ignored because the presence of clusters of fine granular microcalcifications is an important sign for intraductal carcinoma *in situ*.

The binary images resulting from the CAD scheme are used as input to the clustering algorithm detailed in the next chapter. This determines the location of all potential microcalcification clusters.

Chapter 7

Locating clusters of microcalcifications

7.1 Introduction

Clustered microcalcifications are an important early indicator of malignancy and are sometimes the only indication of breast cancer visible on a mammogram. The presence of individual (isolated) microcalcifications are in most cases not clinically significant. Therefore, it is important that a computer detection scheme distinguish between clustered and individual microcalcifications [116]. Figure 7.1 provides examples of three mammogram subimages each containing a single microcalcification cluster. The properties of such clusters must be understood in order to determine the best method of cluster detection.

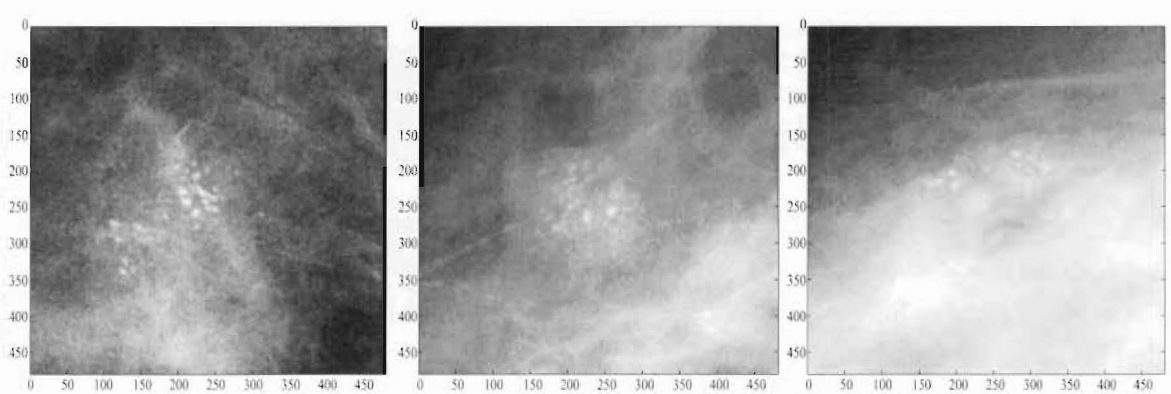


Figure 7.1 Three mammogram subimages from FSD-B each containing a single microcalcification cluster.

The computer aided detection scheme described in Chapter 6 produces a binary image identifying potential individual microcalcification objects. In addition to genuine microcalcifications, falsely identified objects commonly associated with curvilinear structures or image noise may be detected. At the completion of processing with the microcalcification CAD scheme, artifacts may appear in addition to the successfully detected microcalcifications. Fortunately, artifacts tend to occur at random locations or along curvilinear structure boundaries making them distinguishable from genuine microcalcification clusters.

This chapter describes in detail a new hybrid method for achieving the segmentation of microcalcifications into those which form part of a cluster and those that do not. The method combines an algorithm which has formerly been proposed in the literature (mode clustering [39, pp. 127-128]), with a refinement algorithm based on Voronoi sets [147]. The clustering scheme is designed to detect true microcalcification clusters while ignoring isolated individual microcalcifications and falsely identified signals. It should be noted that the clusters determined in this process are not classified into benign or malignant cluster types.

Section 7.2 begins with an outline of the existing microcalcification clustering techniques described in the literature. Pre-processing for application of the new hybrid clustering scheme is described in Section 7.3. Section 7.4 reviews various general clustering techniques identifying their advantages and disadvantages for application to microcalcification cluster detection. An outline of the new hybrid clustering method is given in Section 7.5. The first stage of this scheme (mode clustering) is described extensively in Section 7.6. Section 7.7 details the concept behind Voronoi neighbours laying the groundwork for the second stage of the new hybrid scheme discussed in Section 7.8. Finally, Section 7.9 provides the results of a study of the new clustering technique using images from FSD-B containing a variety of microcalcification cluster patterns.

7.2 Existing methods

The grouping of detected microcalcifications into clusters has not been extensively researched in the literature. Common to all methods used is the initial transformation from a binary image containing a small connected region ('blob') for each detected microcalcification into a set of representative points, one for each microcalcification. Nishikawa *et al* [117] describe in detail a recursive area-point transformation for this purpose. Since many image processing software packages, including MATLAB, have pre-programmed functions to label connected regions in a binary image and the algorithms have been reported elsewhere [57, 101], this part of the algorithm is not considered further. It is important to note however that all such methods are relatively computationally expensive and the area-point transformation may in fact take more time than the rest of the clustering process.

Having transformed the original segmented image into a set of representative points, the methods reported in the literature attempt to group them in a way that is consistent with the perception of an expert radiologist. Chan *et al* [26] describe a *growing* technique by which individual points that are within a particular radius of an existing cluster centre are incorporated into the cluster. Their method is not dissimilar to the mode clustering method described in Section 7.6 and used as part of the new hybrid clustering algorithm. It has been later reported that Chan *et al*'s method has a high rate of false cluster detections [117]. The method is highly dependent on determining the exact spatial location of all potential microcalcifications in the mammogram. Additionally, microcalcifications are sometimes grouped incorrectly, particularly when they cover a large area compared to the clustering radius.

Karssemeijer and Veldkamp [71,150] constructed a set of disks, each centred on one of the microcalcification location points. A group of disks which touch or overlap are considered to be a single cluster. Although the technique is very simple to implement, its performance is quite tightly coupled to the choice of disk diameter and it tends to falsely detect points associated with CLS as clusters (i.e. several artifacts arranged along a line-like structure).

Nishikawa *et al* [117] report a method for cluster detection which passes a kernel of predefined size and shape over the area of the image. If three or more detected points lie within the kernel for any kernel location, all of those points are taken to be part of a cluster. Thus the output of the method is an estimate of the set of clustered microcalcifications. In order to determine which cluster each point belongs to, note is taken of whether the kernels used to select them are overlapped. Finally, clusters whose perimeters are closer than 5mm are combined into a single cluster. It is noteworthy that the authors show a square kernel in their depiction of the algorithm, yet clearly a circular kernel would avoid the performance of the algorithm being dependent on the arbitrary orientation of a cluster within the image. The authors report experimenting with a number of kernel sizes, but it is likely that no one kernel size would achieve optimal cluster detection over a range of mammograms.

7.3 Pre-processing

As mentioned in the previous section, development of a computationally efficient clustering scheme requires each microcalcification to be represented as a single point in image space. Figure 7.2(a) shows part of a digitised film-screen mammogram image containing a microcalcification cluster. Figure 7.2(b) illustrates the binary image output from the CAD scheme described in Chapter 6. The sensitivity of the CAD scheme to microcalcifications has been increased from that reported in Chapter 6 to increase the number of potential microcalcification objects detected. This is done to better illustrate the clustering process in this chapter.

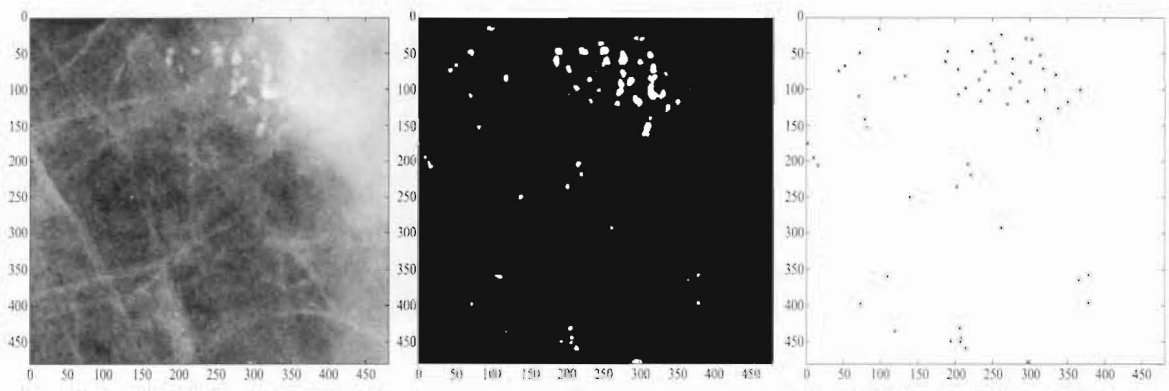


Figure 7.2 (a) Original mammogram subimage (b) Segmented microcalcification pixels from CAD scheme. (c) Single point representation of each microcalcification.

The *bwlabel* function in MATLAB [101] has been used to label each connected region in the image in Figure 7.2(b) with the unique index, $i = 1, 2, \ldots$. The centroid (x_i', y_i') of the ith microcalcification object is used to represent its location in the work reported here (see Figure 7.2(c)). Knowledge of the pixels associated with each microcalcification object enables the centroid to be calculated as

$$x_i' = \frac{1}{N} \sum_{j=1}^{N} x_j, \qquad i = 1, 2, 3, \ldots, \tag{7.1}$$

and

$$y_i' = \frac{1}{N} \sum_{j=1}^{N} y_j, \qquad i = 1, 2, 3, \ldots, \tag{7.2}$$

where N is the total number of pixels in the microcalcification being analysed. The centroids of the microcalcification objects are henceforth referred to as the *points*.

7.4 Review of clustering methods

Data clustering or simply *clustering* has been an active research topic for a number of years. In a recent comprehensive review, Jain *et al* [66] trace its origins back to the early 1970's and cite the application of clustering in several areas, including pattern classification, image segmentation and document retrieval. Some of the techniques used have a form of training built in, and many of the popular and successful algorithms are based on the assumption that all points should be allocated to a cluster. This assumption seems inappropriate in the mammographic application.

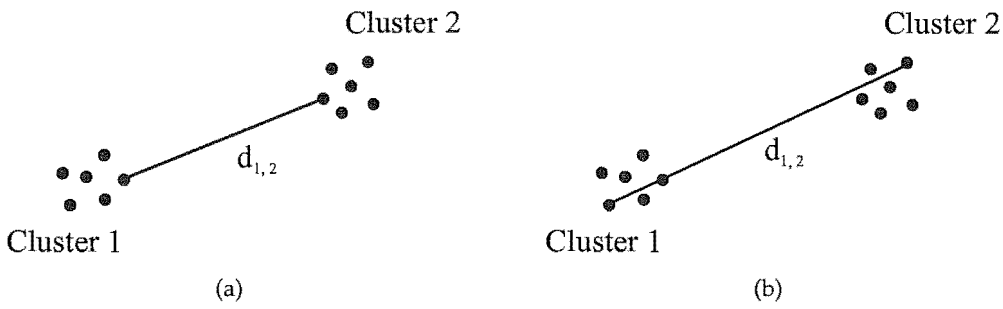


Figure 7.3 (a) Single linkage distance $d_{1,2}$ between cluster 1 and 2. (b) Complete linkage distance $d_{1,2}$ between cluster 1 and 2.

All clustering methods rely on some form of similarity measure between the points. The simplest such measure and the seemingly most appropriate in this application is the Euclidian distance defined as

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \tag{7.3}$$

where i and j index the two points being analysed. In the following, *distance* should be assumed to be the Euclidian distance in 2 dimensions.

Jain *et al* [66] categorise clustering methods into *hierarchical* and *partitional* types. Agglomerative hierarchical techniques apply a series of fusions which may run from n clusters containing a single point to a single cluster containing all points. Partitional techniques on the other hand seek to optimise a cost function by partitioning points into a predefined number of clusters.

7.4.1 Hierarchical techniques

The most well known hierarchical techniques are the single-link and complete-link techniques [39, pp. 55-61]. These two agglomerative methods only differ in the way they characterise the similarity between a pair of clusters. In the single-link method, the distance between two clusters is the minimum between any pair of points, one in each cluster (see Figure 7.3(a)). In the complete-link method, the distance between two clusters is the maximum between any pair of points, one in each cluster (see Figure 7.3(b)). Both cases start with n clusters each containing a single point. The closest pair of clusters are found and merged at each step. This process is repeated until the maximum distance criteria is met.

While these techniques are extremely simple to implement and relatively efficient for moderately sized sets of points, both possess properties that are inappropriate for microcalcification cluster detection. The complete-link algorithm produces tightly bound or compact

clusters. If a cluster is well dispersed the method is unlikely to detect the entire cluster. The single-link algorithm, by contrast, suffers from a chaining effect. It has no obvious stopping criterion (other than eventually stopping when all points have been placed into a single cluster) and tends to produce clusters that are straggly or elongated.

7.4.2 Partitional techniques

The best known example of a partitional clustering technique is the K-means algorithm [66]. This method proceeds by firstly specifying a predefined number of cluster centres K and randomly defining their initial positions. Each point is assigned to the closest cluster centre based on its Euclidian distance from each centre. Recomputation of the cluster centres is then determined by minimising the mean squared error between the current cluster memberships and the cluster centres. The process is repeated until convergence occurs (i.e. there is no reassignment of points from one cluster to another, or the mean squared error ceases to decrease significantly from one iteration to the next). The K-means algorithm is popular because it is easy to implement, and its time complexity is low.

A major problem accompanying the use of partitional clustering techniques is the requirement to know the number of clusters. Such knowledge is clearly unavailable in the application of screening mammograms for the presence of microcalcification clusters. Additionally, many variants of partitional clustering assume that all points must be assigned to a cluster, which is yet again unsuitable for microcalcification cluster detection.

A method classified by Jain *et al* as partitional yet one that clearly shows some of the properties of the hierarchical techniques is mode clustering [39, pp. 127-128]. Since this has been chosen to form the first stage of the hybrid clustering method presented in this chapter, it is described in more detail in Section 7.6.

7.5 New hybrid clustering method

The shortcomings of both the standard available data clustering methods and the reported microcalcification clustering techniques has motivated the development of a new hybrid method. Observation of a number of microcalcification clusters in mammograms identified by expert radiologists suggest two particular characteristics which are distinctly different from most clustering tasks:

1. The points in a cluster are distributed approximately uniformly throughout the cluster. In many other data clustering problems the model for the data is that of a statistical distribution of points about a mean location for each cluster that might be described as a *Gaussian cloud* of points. Many standard clustering methods, especially the partitional methods, are based on this model.

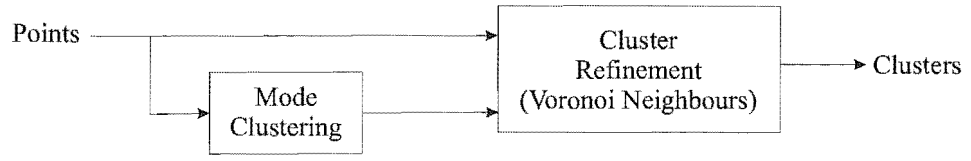


Figure 7.4 Implementation of the new hybrid clustering scheme.

2. There is no prior expectation that all points lie within clusters. In fact there is a high probability that no clusters are present (if the set of mammograms is from the general population). Care must be taken, therefore, not to force points into clusters.

In addition, observation indicates that two or more clusters in a single mammogram may have quite distinctly different densities of points, suggesting that the clustering method must adapt to the average spacing of points within each cluster.

The new hybrid clustering scheme has been developed with these characteristics in mind [16]. It comprises two stages as depicted in Figure 7.4:

1. Mode clustering
2. Cluster refinement based on Voronoi neighbours.

Each stage of the hybrid method is now described in detail.

7.6 Mode clustering

By depicting individual microcalcifications as points in image space, a natural concept of clustering suggests that there should be parts of the space in which the points are very dense, separated by parts with low density. Mode clustering attempts to search for the regions of high density in the data, where each such region is taken to signify a different cluster. This method is a derivative of single-link clustering, but attempts to overcome the chaining effect which is one of the main problems with the single-link technique.

The first step in mode cluster analysis is the calculation of a distance matrix D. This matrix is symmetric ($d_{i,j} = d_{j,i}$) and contains all point-to-point Euclidean distances. A search is made by considering a radius r surrounding each point in image space and counting the number of points falling within this radius. A point is defined to be *dense* if it has at least M neighbouring points the distances to which are less than or equal to r. The algorithm proceeds as follows:

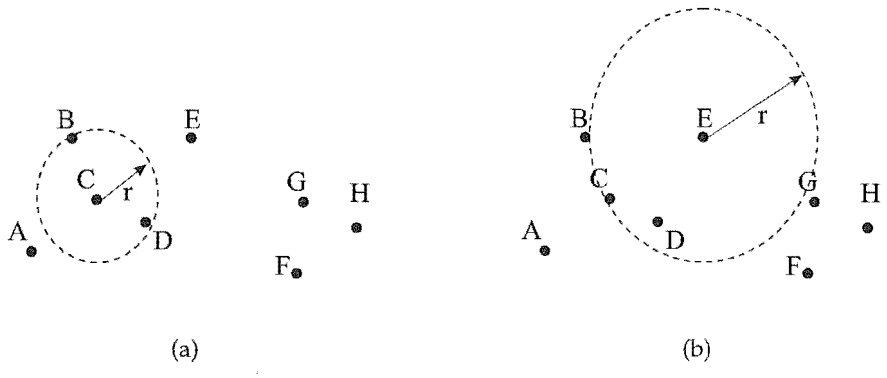


Figure 7.5 (a) Point C is classified as dense at the radius shown with $M = 2$. (b) Point E is classified as dense at the radius shown with $M = 2$.

1. Radius r is initialised to the minimum distance between any pair of points r_{min}.
2. Radius r is gradually increased until it reaches r_{max}.
3. At each increase in r, any new dense points are processed according to 4, 5, or 6 below.
4. If the new dense point has distances $> r$ from all points that have already been allocated to a cluster, a new cluster is initiated.
5. If points with distances $\leq r$ from the new dense point include one or more points belonging to a single cluster, the new dense point joins that cluster.
6. If points with distances $\leq r$ from the new dense point belong to more than one existing cluster, all clusters represented are joined into a single cluster including the new dense point.
7. Clustering stops when the radius r reaches r_{max}.

Figure 7.5 illustrates the classification of a point as *dense* when the value $M = 2$. The point C is classified as dense for a small radius r (Figure 7.5(a)), while point E is not classified as dense until the radius reaches a much higher value (Figure 7.5(b)).

Figure 7.6 shows the final clusters determined after application of the mode clustering algorithm to the points given in Figure 7.5(a) (with $M = 2$). The value of r_{max} is chosen to coincide with the radius shown in Figure 7.5(b). The diagram in Figure 7.6(b) illustrates the fusion of points into clusters at each successive stage of the analysis as r is increased. The diagram reads from bottom to top. The first point to be classified as dense is point C, followed by points F, G and H and so on.

The advantages of the mode clustering method over other clustering techniques are that: 1) it retains the simplicity of the hierarchical techniques (no calculations are performed after the initial distance matrix is formed, just comparisons); and 2) it has a simple stopping

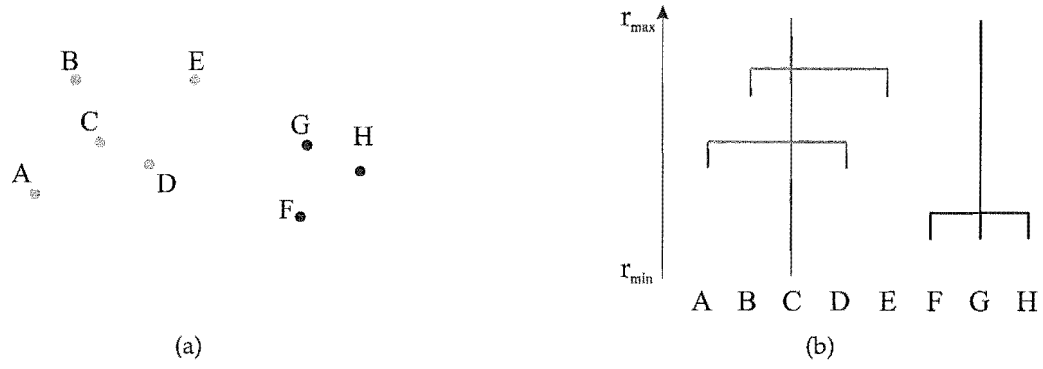


Figure 7.6 (a) Final clustering of points when using mode cluster analysis. (b) Linkage of points for different values of r.

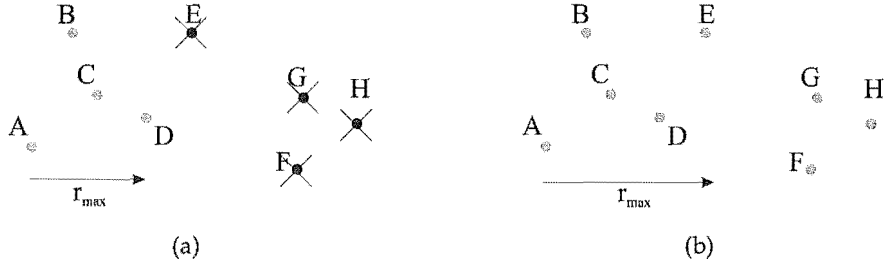


Figure 7.7 (a) Final clustering of points for $M = 3$. Points marked with a '×' are not assigned to any cluster. (b) Final clustering of points for $M = 3$ with r_{max} increased.

criterion. The disadvantages are: 1) since it purely agglomerates points it can produce *unnatural* clusters which are clearly incorrect to the human observer; and 2) it is quite sensitive to the choice of the two parameters, M and r_{max}.

Figure 7.7 illustrates how changes in the parameters M and r_{max} can produce very different clustering results. Figure 7.7(a) shows the single cluster formed by mode clustering when $M = 3$ and r_{max} has the value indicated by the length of the arrow shown. Points marked with a '×' have not been assigned to any cluster. Figure 7.7(b) shows the clusters found when $M = 3$ and r_{max} is increased slightly. All points are now considered to be part of a single large cluster. These results illustrate the high sensitivity of the mode clustering algorithm to the choice of parameter values, even for a simple example such as this.

On account of the disadvantages of the mode clustering technique, this algorithm is employed only as the first stage in the new hybrid method presented here, choosing M and r_{max} conservatively and relying on the second stage (described in Sections 7.7 and 7.8) to

tidy up the clustering. The second stage incorporates measures based on Voronoi neighbours [147].

When choosing the values for M and r_{max} consideration of the expected geometry of a cluster is required. If the value of M is chosen too high (> 3) points on the periphery of a cluster are likely to be excluded. On the other hand choosing M too small ($M = 1$) can result in straggly or elongated clusters. The value determined for r_{max} must be large enough to cover multiple points within a true cluster however should not been chosen too high that it groups together points that are highly dispersed or relatively isolated. The exact values chosen for these two parameters are given in Section 7.9.

7.7 Cluster measures based on Voronoi neighbours

The purpose of the second stage in the hybrid clustering scheme is to refine the clusters already detected by the mode clustering algorithm, ensuring the correct detection of all points within each cluster. This refinement includes the merging of two or more clusters if their distance from one another is comparatively small, and the addition of individual points to a cluster based on the linkage (described below) between an individual point and a cluster [16].

In order to exploit the uniform distribution of points (identified in Section 7.5) appropriate measures are required. Traditionally standard statistical parameters such as the root mean squared point to cluster centre distance and the root mean squared point to point distance have been used in assessing whether a point should join a cluster [39, pp. 37-53, 66]. These well known techniques however do not provide a measure of uniformity within a cluster which is important for application to microcalcification cluster detection.

Consider the two clusters of points depicted in Figure 7.8. The mean distance between all points within the cluster and the cluster centre are identical in both cases. However the nature of the two clusters is clearly different and the decision of whether an additional point should be included in either cluster should reflect that difference. The concept of the *Voronoi neighbours* is introduced in this section in order to develop a set of useful measures.

7.7.1 Voronoi neighbours

The Voronoi diagram is well known in the areas of graph theory and discrete mathematics [1,147]. Let $p_1, p_2, \ldots, p_n$ denote n distinct points in the image plane. The Voronoi diagram partitions the plane into n disjoint regions or tiles such that the tile corresponding to p_i is

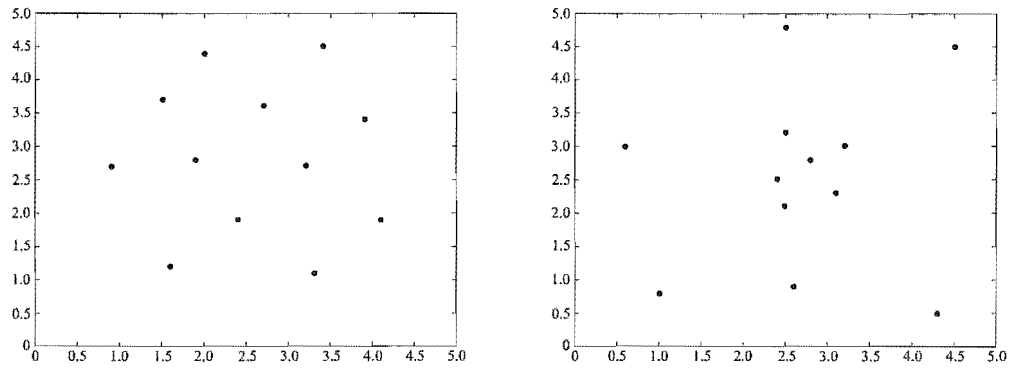


Figure 7.8 (a) Approximately *uniform* distribution of points. (b) *Gaussian cloud* distribution of points.

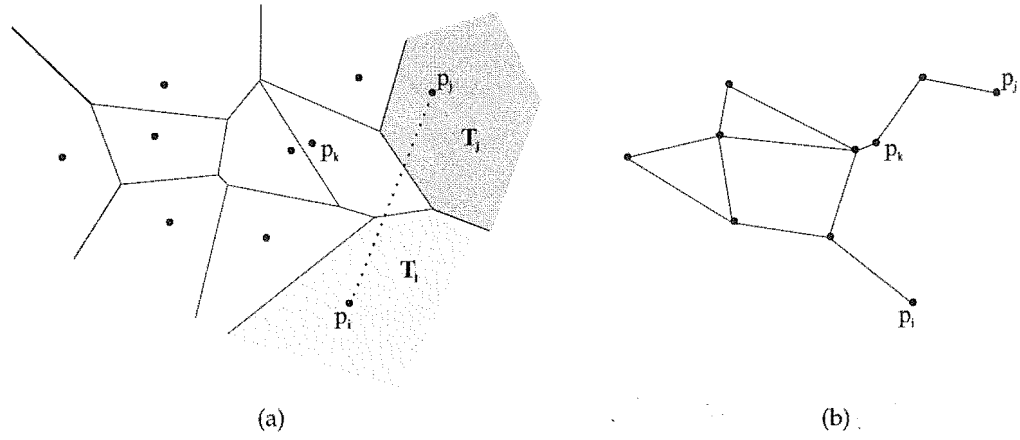


Figure 7.9 (a) Voronoi diagram showing the disjoint tiles surrounding each point. (b) The same set of points as in (a) with Voronoi neighbours connected by lines. Note that p_i and p_j are not Voronoi neighbours.

the set of locations in the plane T_i defined by

$$T_i = \{\chi : d(\chi, p_i) < d(\chi, p_j) \quad \forall \quad j \neq i\}, \qquad i = 1, 2, \ldots, n, \tag{7.4}$$

where χ is any location in the plane and d denotes the Euclidian distance. An example of a Voronoi diagram is given in Figure 7.9(a).

The ith and jth points, p_i and p_j, are Voronoi neighbours if the location on the straight line joining p_i and p_j which is equidistant from p_i and p_j lies on the junction between tiles T_i and T_j. In other words p_i and p_j are Voronoi neighbours if the midpoint of the straight line joining them is not closer to any other point in the total set. Although the points p_i and p_j shown in Figure 7.9(a) have adjacent tiles T_i and T_j, they are not Voronoi neighbours because the midpoint of the straight line joining them is closer to point p_k. Figure 7.9(b)

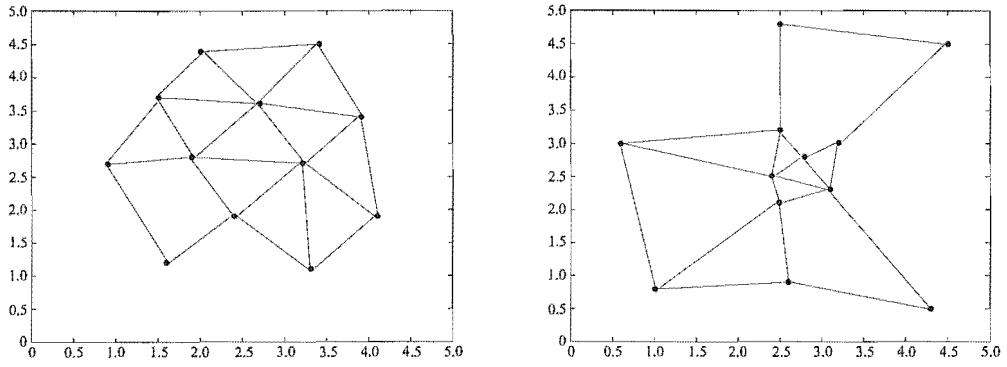


Figure 7.10 Voronoi neighbours connected by a straight line for (a) an approximately *uniform* distribution of points and (b) a *Gaussian cloud* of points.

shows the Voronoi neighbours associated with the Voronoi diagram in Figure 7.9(a). The significance of determining the Voronoi neighbours is that the distances between them tend to be equal for a uniform distribution of points. Several measures are defined on the basis of Voronoi neighbours in the next subsection.

7.7.2 Voronoi distance measures

As mentioned previously most standard statistical measures of scatter do not provide a measure of uniformity within a cluster. For this reason a new set of distance measures based on Voronoi neighbours is defined here [16].

First the *Voronoi distance*, v_{ij}, is denoted as the Euclidean distance between the Voronoi neighbours indexed by i and j. If p_i and p_j are not Voronoi neighbours, $v_{ij} = \infty$. Figures 7.10(a) and (b) show the Voronoi neighbours for the distribution of points given in Figures 7.8(a) and (b) respectively. The maximum finite Voronoi distance for the approximately uniform distribution of points (Figure 7.10(a)) is 1.66 while the minimum distance is 0.86. For the Gaussian distribution however (Figure 7.10(b)) the maximum finite Voronoi distance is 2.24 and the minimum 0.41. This large difference in distance variation between the two distributions indicates that certain measures based on $v_{i,j}$ can be used to provide a measure of uniformity within a cluster.

Within the kth cluster, the *mean Voronoi distance*, ${}^k v$, is defined to be the arithmetic mean of all the finite Voronoi distances occurring between elements of the cluster. For example in Figure 7.10(a) the approximately uniform distribution of points has ${}^k v = 1.184$ and in Figure 7.10(b) the Gaussian distribution has ${}^k v = 1.277$, which are very similar in value.

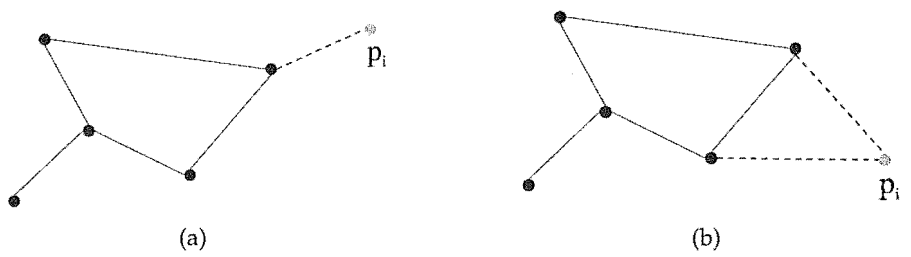


Figure 7.11 (a) External point p_i with one Voronoi neighbour. (b) External point p_i with two Voronoi neighbours

The *Voronoi distance deviation* for the kth cluster, ${}^{k}\sigma$, is defined as the standard deviation of all the finite Voronoi distances occurring between elements of the cluster. This is a measure of uniformity of point locations within the cluster. For example, in Figure 7.10(a) the uniform distribution of points has ${}^{k}\sigma = 0.199$ indicating a very small standard deviation of finite Voronoi distances. However, in Figure 7.10(b) the Gaussian distribution has ${}^{k}\sigma = 0.686$ showing a much larger standard deviation and highlighting the non-uniformity of the cluster.

The *mean point-to-cluster Voronoi distance* from an unclustered point p_i to the kth cluster, ${}^{k}v_i$, is defined to be the arithmetic mean of all of the finite Voronoi distances occurring between p_i and the elements of the cluster. Should p_i not have any Voronoi neighbours in the kth cluster, ${}^{k}v_i = \infty$. The *mean cluster-to-cluster Voronoi distance*, ${}^{km}v$, for clusters k and m, is defined to be the arithmetic mean of all the finite Voronoi distances occurring between the elements in cluster k and the elements in cluster m. If no Voronoi neighbours exist between the two clusters ${}^{km}v = \infty$.

The distance measures defined above are useful for deciding how similar the distance between a point and its nearest neighbour or neighbours in a cluster is to the density within the cluster. However, this is not sufficient. Consider the situation depicted in Figure 7.11. Both external points p_i in Figures 7.11(a) and (b) are similar distances from an already defined cluster. It seems more likely that the external point p_i in Figure 7.11(b), with two Voronoi neighbours in the cluster, belongs as part of the cluster compared to the external point in Figure 7.11(a), which has only one Voronoi neighbour in the cluster. For this reason an additional quantity known as the *Voronoi linkage order* ${}^{k}\lambda_i$ is defined between the kth cluster and an unclustered point p_i. The Voronoi linkage order corresponds to the number of Voronoi neighbours to p_i in the cluster. The boundary limit of whether an external point should be included in a nearby cluster is related to the Voronoi linkage order as explained in the next section.

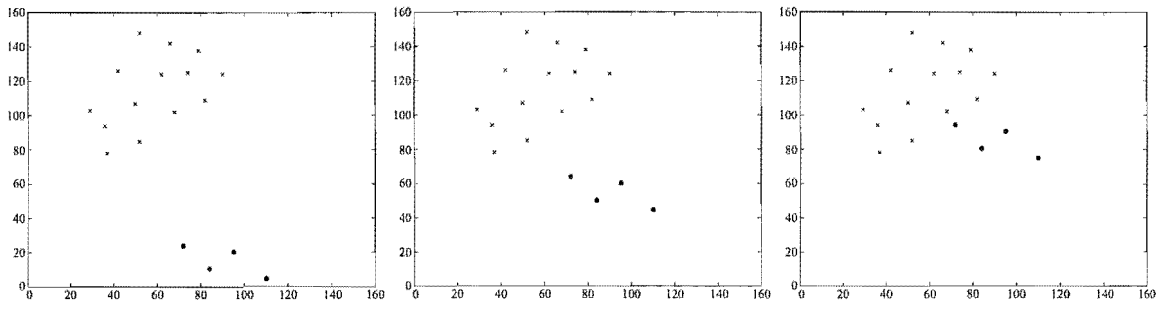


Figure 7.12 (a) Two highly separated clusters with $\alpha = 3.38$ in Equation 7.5. (b) Two clusters positioned such that $\alpha = 2.25$ in Equation 7.5. (c) Two clusters placed close together with $\alpha = 0.95$ in Equation 7.5.

7.8 Clustering refinement

A set of *ad hoc* rules have been established for applying the measures, defined in the previous section, to refining the preliminary clustering generated by the mode clustering algorithm [16]. The rules are:

1. Two clusters k and m should merge if ${}^{km}v \leq \alpha.({}^{k}v + {}^{m}v)/2$, where α is an empirically determined positive constant.
2. An unclustered point p_i should join the kth cluster if ${}^{k}v_i \leq {}^{k}v.(1 + \beta.{}^{k}\lambda_i) + \gamma.{}^{k}\sigma$, where β and γ are empirically determined positive constants.
3. Clusters must have three or more points.

The converse of rule 2 is that a point on the periphery of a cluster should leave the cluster if the condition is not satisfied.

Refinement rule 1 states that the mean cluster to cluster Voronoi distance should be smaller than an amount proportional to the average of the mean Voronoi distances for each individual cluster before considering merging. To illustrate the effect of the choice of α, consider the two clusters depicted in Figure 7.12. The value of α satisfying the equality of refinement rule 1,

$$ {}^{km}v = \alpha.({}^{k}v + {}^{m}v)/2, \tag{7.5} $$

was calculated as the clusters were moved progressively towards each other.

Figure 7.12(a) shows two highly separated clusters. From visual inspection it is obvious these clusters are well separated and therefore should not merge into a single cluster. In this case a value of $\alpha = 3.38$ satisfies Equation 7.5. This value is much too high for the

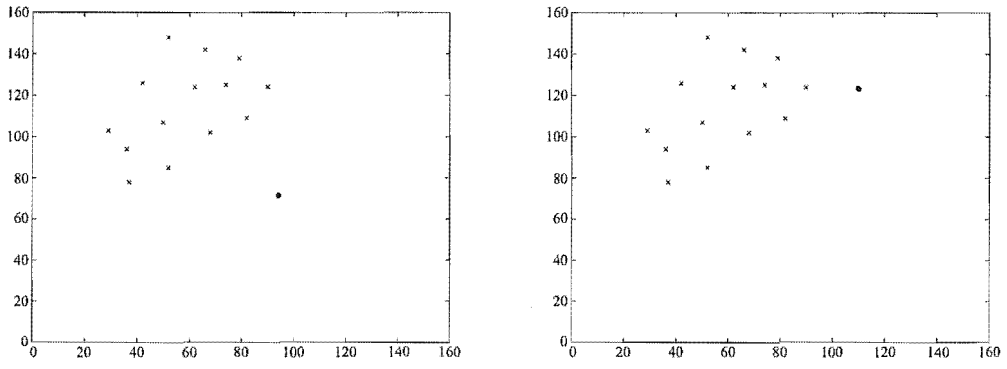


Figure 7.13 (a) An unclustered point p_i and a cluster k with ${}^k\lambda_i = 3$. (b) An unclustered point p_i and a cluster k with ${}^k\lambda_i = 1$.

application of microcalcification cluster detection in which clusters are considered to be relatively uniformly distributed. The clusters shown in Figure 7.12(b) are positioned closer together and appear to warrant a decision to merge. The value $\alpha = 2.25$ satisfies Equation 7.5 in this case. Finally, Figure 7.12(c) show the two clusters positioned closer together again where a merge is definitely the appropriate action ($\alpha = 0.95$ in Equation 7.5). The value $\alpha = 2.25$ was chosen for the study presented in Section 7.9.

Refinement rule 2 is established to determine if an unclustered point should join an already defined cluster. This rule has two empirically determined positive constants β and γ each having a different effect. The γ constant is associated with the Voronoi distance deviation, ${}^k\sigma$. A uniform cluster has very low ${}^k\sigma$ therefore adding an unclustered point to this type of distribution requires the point to be positioned comparatively close to the periphery of the cluster ensuring the mean point-to-cluster Voronoi distance, ${}^kv_i$, is low. However, for a less regular cluster with higher ${}^k\sigma$, a point further from the periphery of the cluster may be acceptable. The value of γ chosen for the study presented in Section 7.9 was 1.

The β constant in rule 2 is associated with the Voronoi linkage parameter, ${}^k\lambda_i$. This term is designed to ensure an unclustered point p_i is more likely joined to an already defined cluster k when the linkage between the cluster and that point is high. Figure 7.13 provides two illustrative examples of refinement rule 2 when $\beta = 0.5$ and $\gamma = 1$. The position of the unclustered point in both cases satisfies

$$ {}^kv_i =^k v.(1 + \beta.{}^k\lambda_i) + \gamma.{}^k\sigma. \tag{7.6}$$

In Figure 7.13(a) the unclustered point p_i has a Voronoi linkage order ${}^k\lambda_i = 3$, therefore the point is able to be placed further from cluster k while still satisfying refinement rule 2. However, in Figure 7.13(b) p_i has a Voronoi linkage order ${}^k\lambda_i = 1$. This means the point

must be placed closer to cluster k in order to satisfy refinement rule 2. The value of 0.5 was chosen for β in the study in Section 7.9.

After determining the most appropriate values for the positive constants associated with each rule the refinement stage proceeds by applying an iterative process successively as follows:

1. Apply rule 1 to all clusters.
2. Apply rule 2 to all unclustered points p_i in combination with any cluster k for which ${}^k\lambda_i$ is non- zero.
3. Apply rule 3 to all clusters.
4. Apply rule 2 (converse) to all points in each cluster (ignoring the effect on the cluster measures ${}^kv_i$, ${}^kv$ and ${}^k\sigma$ removing that point would make).

The iteration proceeds until no further reallocation of points or clusters is made.

7.9 Results

The clustering scheme presented in this chapter was applied to all images in FSD-B (see Section 2.2). This database contains 50 mammogram images with a total of 37 malignant microcalcification clusters (no benign microcalcification clusters). The images containing microcalcification clusters have an associated *overlay* image which provides annotations marking the location and size of each cluster. It was found that the annotations (which were provided by expert radiologists) were generous in size. For this reason a cluster detected by the new hybrid method is classed as a true positive (TP) if all microcalcifications associated with the cluster are within the marked area on the overlay image. All other clusters found by the algorithm are considered to be false positives (FP). The criteria for a TP detection is very restricting and ensures the clusters found are accurately localised. No verification of the detection of individual microcalcifications is performed. A more detailed description of the detection criteria for this particular database is given in Section 8.5.

The parameters for both the mode clustering (stage 1) and the Voronoi refinement (stage 2) in the new hybrid clustering scheme were assigned the values:

$$\begin{aligned} &\text{Stage 1:} \quad r_{max} = 30 \quad M = 3 \\ &\text{Stage 2:} \quad \alpha = 2.25 \quad \beta = 0.5 \quad \gamma = 1, \end{aligned}$$

for a 50μm resolution image. It should be noted that r_{max} is the only parameter dependent on the resolution of the mammogram.

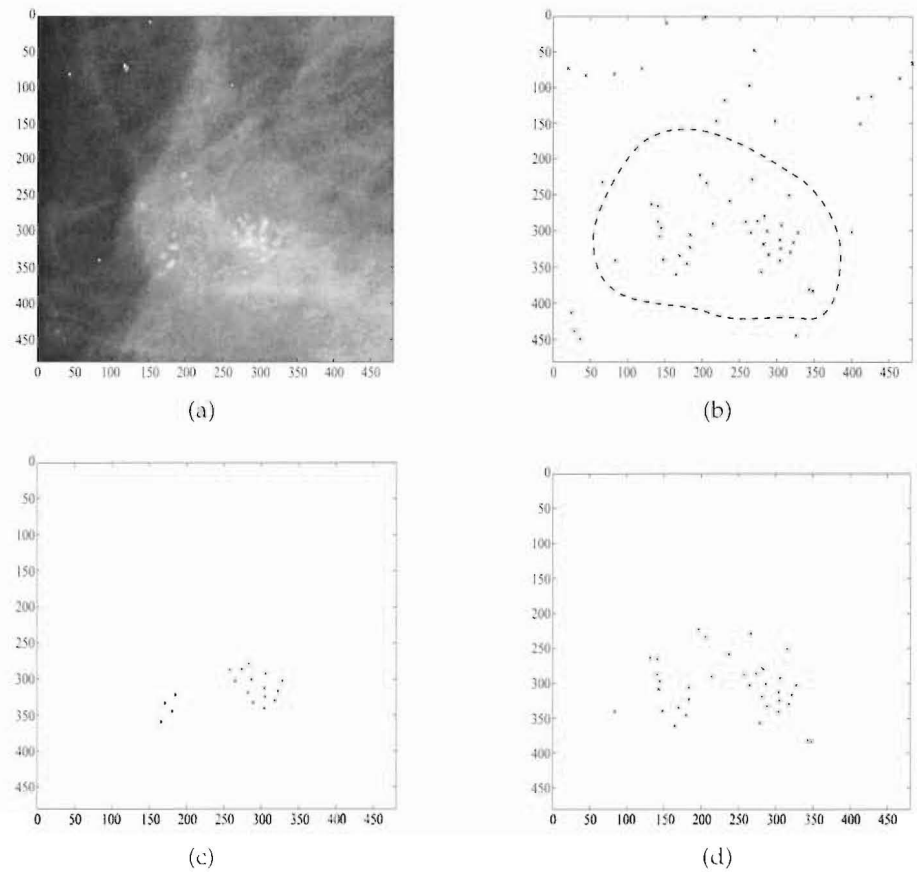


Figure 7.14 (a) Subimage from FSD-B containing one true microcalcification cluster. (b) Potential microcalcification points identified by the CAD scheme. (c) Clusters detected after mode clustering. (d) Clusters detected after applying the Voronoi refinement stage.

Figures 7.14 to 7.16 provide illustrative examples of the typical operation of the clustering scheme on three subimages from FSD-B. All of the selected subimages contain at least one microcalcification cluster (subfigures (a)). Subfigures (b) show the complete data set of points detected from the CAD scheme, which include both true microcalcifications and artifactual detections. Also shown in these subfigures are the annotation boundaries of the true microcalcification clusters.

The clusters shown in subfigures (c) are those determined after application of the mode clustering algorithm (1st clustering stage). In all cases shown part of each true microcalcification cluster is detected. In Figure 7.14(c) the true cluster is detected as two smaller clusters. In Figure 7.15(c) one false positive cluster is detected.

Subfigures (d) show the final clusters detected after application of the Voronoi refinement algorithm (2nd clustering stage). It is clear from these subfigures that all true microcalcification clusters have been detected in their entirety. The refinement stage has joined the

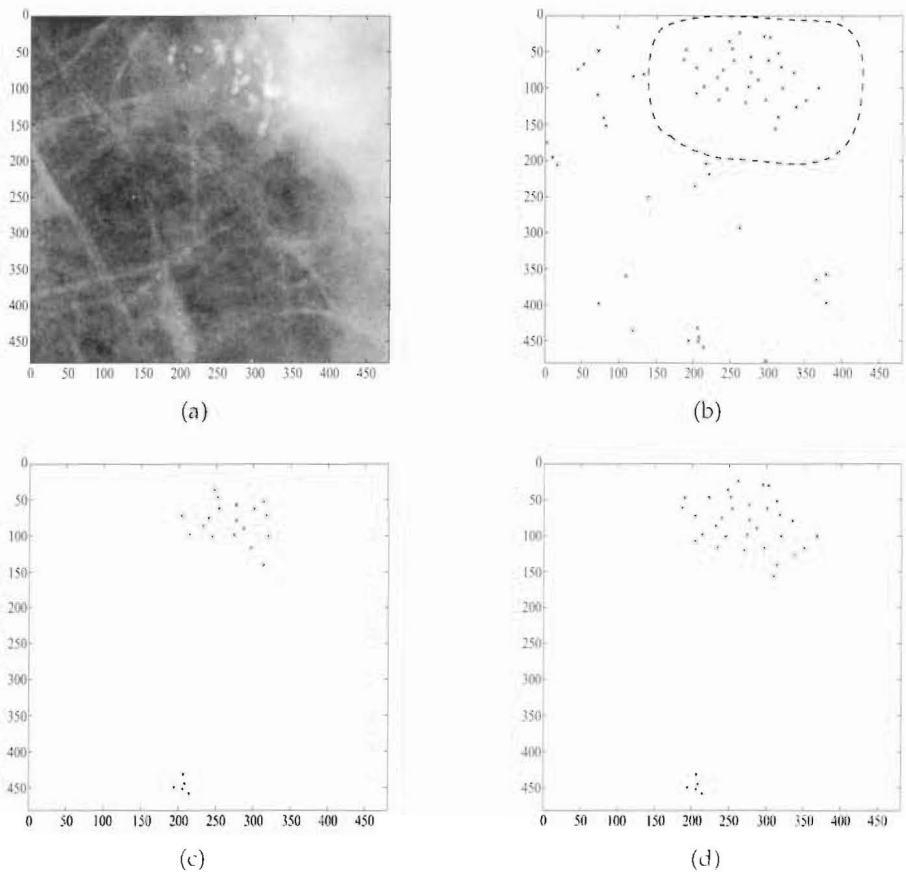


Figure 7.15 (a) Subimage from FSD-B containing one true microcalcification cluster. (b) Potential microcalcification points identified by the CAD scheme. (c) Clusters detected after mode clustering (one TP cluster and one FP cluster). (c) Clusters detected after applying the Voronoi refinement stage (one TP cluster and one FP cluster).

two small clusters in Figure 7.14(c) into one large true positive cluster. The false positive cluster associated with Figure 7.15(a) is still present in the final Voronoi clusters. This is because the layout of the points in this cluster is consistent with the expected properties of a genuine cluster.

After applying the new hybrid clustering algorithm to all images in FSD-B the ability of the scheme was evaluated. In 15 out of 50 cases the mammograms were classified by the radiologist as normal because they did not contain a true microcalcification cluster. Out of these 15 images 3 false positive detections were made. In 30 out of the other 35 cases the algorithm correctly identified genuine clusters, although in one of these a single genuine cluster was identified as more than one. In 28 out of the 30 cases where the true cluster was detected, the Voronoi-distance based refinement method improved the cluster identification by either eliminating false clusters, combining clusters, or by including/excluding points around the periphery of clusters. In the remaining two cases the mode clustering

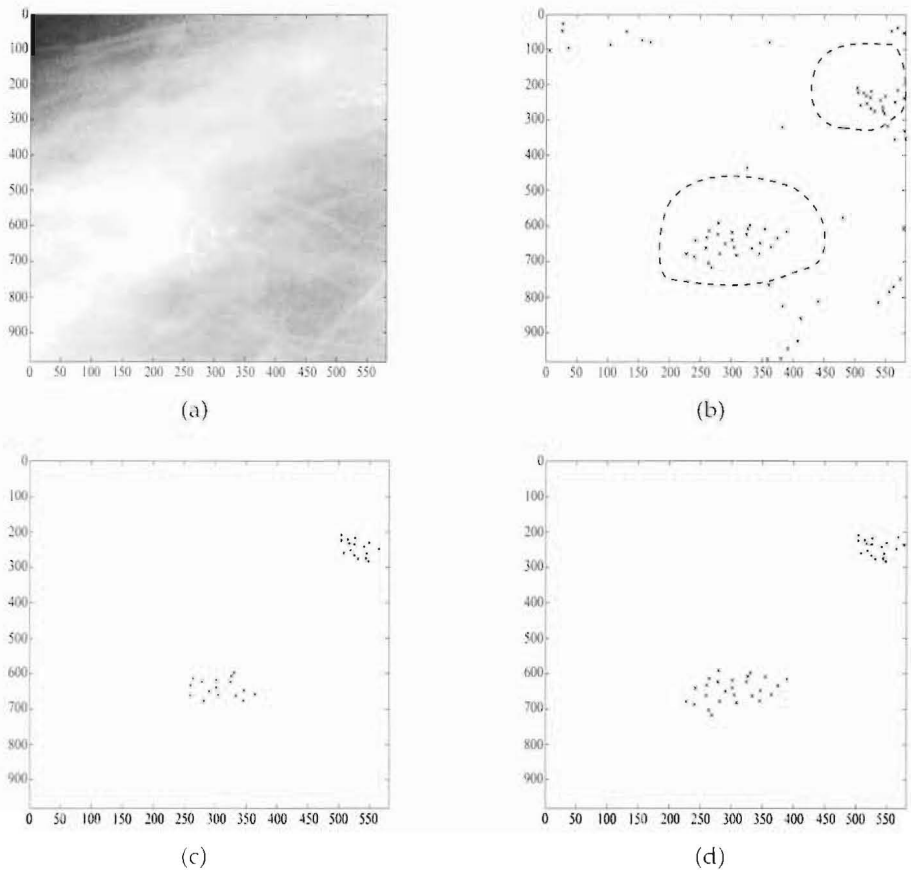


Figure 7.16 (a) Subimage from FSD-B containing two true microcalcification clusters. (b) Potential microcalcification points identified by the CAD scheme. (c) Clusters detected after mode clustering. (c) Clusters detected after applying the Voronoi refinement stage.

had already achieved an optimal cluster detection. In 7 out of the 35 cases, clusters in addition to the genuine ones were incorrectly identified: a total of 9 false positive identifications of clusters were made.

In 5 out of 35 cases a genuine microcalcification cluster was not identified as a true cluster by the hybrid algorithm. This was due to the individual microcalcifications in these clusters being small in number and highly dispersed. Two example microcalcification clusters that went undetected by the hybrid clustering scheme are shown in Figure 7.17. Figures 7.17(a) and (d) show the original mammogram subimages (480×480 pixels). Figures 7.17(b) and (e) respectively illustrate the binary images output from the CAD scheme described in Chapter 6. The single point representation of each microcalcification object used as input to the clustering algorithm is shown in Figures 7.17(c) and (f). It is difficult even for a human observer to determine with certainty that a cluster is present in these figures.

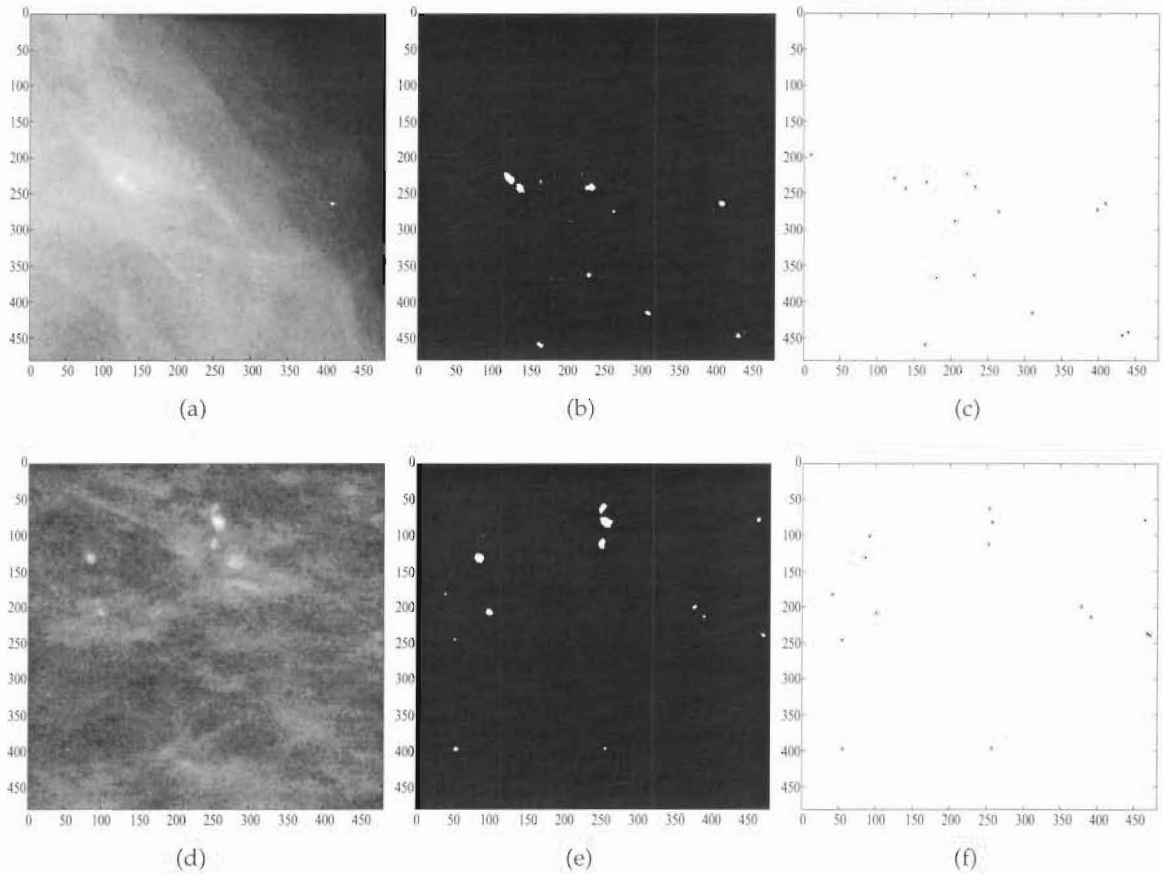


Figure 7.17 (a,d) Subimages from FSD-B each containing a single true microcalcification cluster. (b,e) Segmented microcalcification pixels from the CAD scheme for subimages (a) and (d) respectively. (c,f) Single point representation of each microcalcification object in subimages (b) and (e) respectively.

7.10 Summary

This chapter demonstrated the use of a new hybrid clustering method for finding clusters of microcalcifications within a mammogram image. Given a set of points, each representing a potential microcalcification, the clustering scheme attempts to determine those which form part of a cluster and those that do not.

The method combines a mode clustering algorithm which has formerly been proposed in the literature, with a refinement algorithm based on Voronoi sets. Mode clustering is an effective and efficient method for the first stage of the hybrid scheme however suffers from a high dependency on the parameters chosen. For this reason a second stage employing Voronoi neighbours is applied. Choosing the mode cluster parameters conservatively gives a good starting point for the Voronoi refinement stage. The purpose of the second stage is to refine the clusters already detected by the mode clustering algorithm. This includes

the merging of two or more clusters if their distance from one another is comparatively small, and the addition of individual points to a cluster based on the linkage between an individual point and a cluster.

The results obtained by the new hybrid clustering scheme on the set of mammogram images from FSD-B are encouraging. Genuine microcalcification clusters are not identified by the clustering scheme in only 5 out of 35 cases due to the individual microcalcifications in the cluster being small in number and highly dispersed. Additionally, a few false positive detections are made but in most of these cases the layout of the points are consistent with the expected properties of a genuine cluster.

Chapter 8

CAD system performance evaluation

8.1 Introduction

A new CAD scheme for the detection of microcalcification clusters has been described extensively in Chapters 2 to 7. The scheme is designed to highlight the location of both benign and malignant microcalcification clusters, potentially assisting radiologists in improving the effectiveness of mammography screening. To determine the accuracy of the system and hence the potential contribution to the mammography screening process, a performance evaluation was performed and the results are presented and discussed here. The aim is to optimise system performance based on various algorithm implementations detailed in previous chapters. Establishing the efficacy of a computer aided detection scheme is not a simple task. Strict criteria must be put in place to provide accurate evaluation of the system. Precise interpretation of the information provided by the CAD scheme helps determine the information that can be sought for performance evaluation.

Section 8.2 begins with a general description of the evaluation criteria used for CAD schemes designed to detect abnormalities in medical images. The appropriate method of analysis in this case is the free-response receiver operating characteristic (FROC) analysis, which is outlined in Section 8.3. Section 8.4 provides specific criteria for determining the detection accuracy of a microcalcification CAD scheme. Due to the significantly different image characteristics of the digitised film-screen and direct digital modalities two separate mammogram databases are used, one consisting of digitised film-screen mammograms and the other consisting of direct digital mammograms. Determination of the best criteria for evaluating the CAD scheme applied to each database is described in Section 8.5. Implementation of the FROC analysis for the new CAD scheme is outlined in Section 8.6. Finally, Section 8.7 extensively evaluates the CAD system performance on digitised film-screen mammograms and Section 8.8 details the CAD performance for direct digital mam-

Computer Decision \ True State	IS an abnormality	IS NOT an abnormality
Called an abnormality	True Positive (TP)	False Positive (FP)
Not called an abnormality	False Negative (FN)	True Negative (TN)

Figure 8.1 Possible decisions made during a computer aided detection task. The word *positive* or *negative* corresponds to the decision made, whereas the word *true* or *false* denotes whether or not the decision was correct.

mograms. Results from this in depth study reinforce the findings of previous chapters and determine the best implementation of the CAD scheme for each imaging modality.

8.2 Performance evaluation

When computer algorithms are designed to detect abnormalities in medical images, rigorous evaluation criteria are crucial in assessing the CAD system's performance, since this determines the effect the computer generated information ultimately has on decisions involving patient care.

A decision made during a computer aided detection task falls into one of four possible categories, shown in Figure 8.1 [154]. An image region can be called abnormal (positive) or normal (negative), and a decision can either be correct (true) or incorrect (false). There are two types of errors that can be made: false-negative and false-positive errors. A false-negative (FN) error implies that a true abnormality was not detected. A false-positive (FP) error occurs when a detection corresponds to a normal region, and thus falsely identifies the region as abnormal. Two types of correct decisions can also be made: true-positive (TP) and true-negative decisions (TN). A detection that corresponds to an actual abnormality is called a true-positive. A true-negative decision simply means a normal region was correctly labelled as being normal.

Two measures known as the *sensitivity* and *specificity* of a detection system are related to the decisions identified above as follows: The sensitivity is the rate at which abnormalities are detected and is defined as

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \tag{8.1}$$

The specificity is the rate at which normal regions are correctly labelled as normal and is

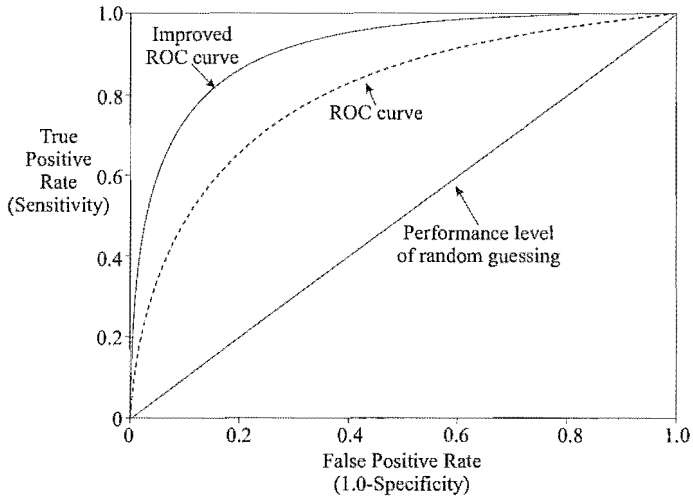


Figure 8.2 Two ROC curves and the expected performance from random guessing.

defined as

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{8.2}$$

The performance of a process that detects abnormalities in medical images can be completely characterised by its sensitivity/specificity tradeoff.

Detection algorithms have parameters that can be varied to alter the TP and FP rates. Each set of parameter values may result in a different (TP,FP) pair, called an operating point. Given the ability of a CAD scheme to determine varying levels of sensitivity, it does not seem sufficient to evaluate or compare different algorithms on the basis of a single operating point. A well accepted method for algorithm evaluation that takes into account the sensitivity/specificity tradeoff characteristics is the receiver operating characteristic (ROC) analysis [23,108,109,154], an example of which is shown in Figure 8.2. The graph presents the FP rate (1.0-specificity) versus the TP rate (sensitivity) for multiple operating points. A maximum threshold for detection (i.e. no detections) corresponds to the lower left point in the ROC curve. As the threshold is lowered the operating point follows a trajectory towards the upper right point.

The area under this curve, usually referred to as the A_z index, is an accepted way of evaluating a systems performance. Perfect diagnostic accuracy means a sensitivity of 100% (or 1.0), a FP rate of 0% and an A_z of 1.0. Random guessing results in an A_z of 0.5.

The ROC is a well known paradigm used for evaluating system performance. In this methodology, it is assumed that for each case presented, a single binary normal/abnormal decision is made and that this is sufficient to characterise detection performance. Unfor-

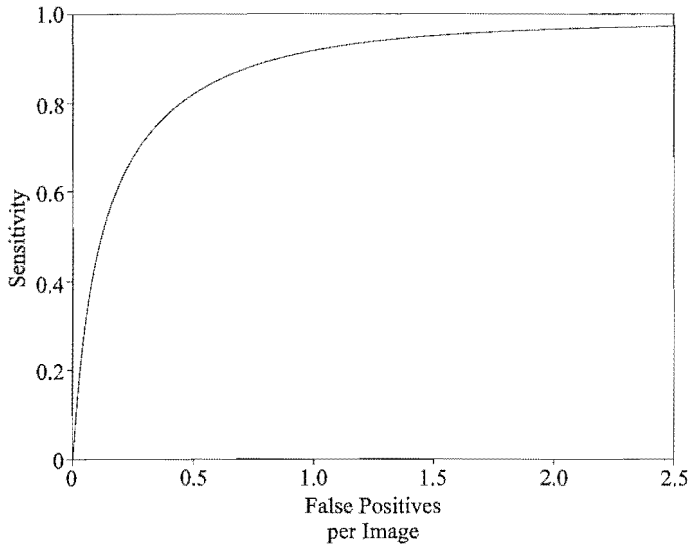


Figure 8.3 FROC curve plotting the sensitivity against the number of false-positives generated per image.

tunately, in a computer aided detection system such as that developed in this thesis there are a number of reasons why the application of ROC analysis is inappropriate. The correct localisation of an abnormality is required for a true-positive detection in a CAD system, however a ROC study does not account for this localisation. Additionally, ROC analysis does not accommodate the presence of multiple lesions on the same image. For an image with two lesions, one CAD scheme may detect only one lesion, while another CAD scheme detects both lesions. In the ROC method of analysis both CAD readings would be scored identically. One final problem with ROC analysis is the required estimation of the false-positive rate. This cannot be determined from a CAD scheme due to the unavailability of the number of TN decisions. For these reasons a free-response receiver operating characteristic (FROC) analysis is used in place of the ROC analysis. This is described in detail in the next section.

8.3 Free response receiver operating characteristic analysis

It is insufficient to simply report the existence of an abnormality in a CAD scheme, because correct localisation of the abnormal area is also required for a true-positive detection. The appropriate method of evaluation in this case is free-response receiver operating characteristic (FROC) analysis [21–23, 154]. FROC analysis permits multiple abnormalities per image and requires correct localisation of abnormalities. A FROC curve is a plot of operating points showing the tradeoff between the sensitivity and the average number of false-positives per image, as shown in Figure 8.3.

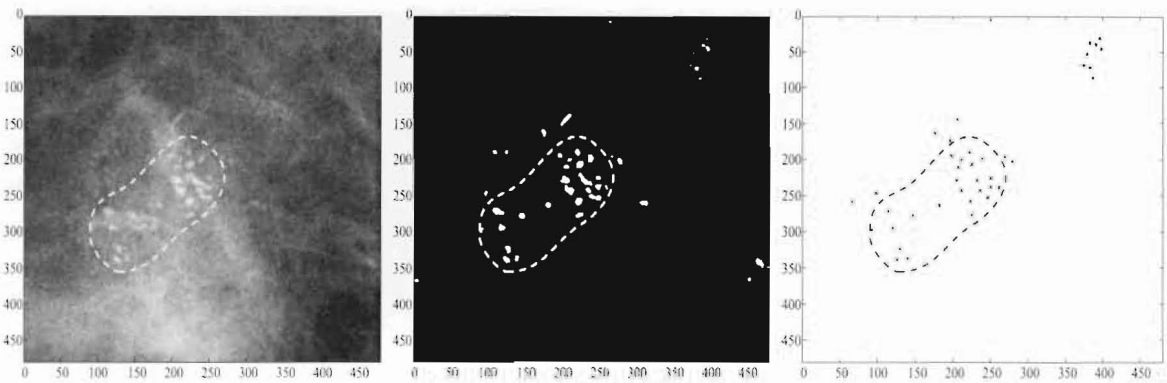


Figure 8.4 (a) Mammogram subimage containing one true microcalcification cluster and ground truth overlay. (b) Binary image output from CAD scheme identifying potential individual microcalcifications. (c) Two clusters detected (one TP and one FP) with ground truth overlay.

The ordinate for a FROC plot is the same as for ROC plots, 0% to 100% sensitivity. The abscissa of the FROC plot begins at zero, but the upper limit is open. Realistically, a detection system is not very useful if too many false-positives are generated, and for most sufficiently large data sets 100% sensitivity is rarely attainable. For these reasons FROC plots rarely show performance above a few false-positives per image. Unfortunately the statistical analysis that can be done with ROC has not been fully developed with FROC [21, 22, 109]. However, the A_z index is still considered a useful measure in evaluating a CAD scheme's performance. The A_z value for a FROC curve, like that of a ROC curve, has a maximum value of 1.0 corresponding to a perfect performance. The A_z value is computed by normalising the area under the FROC curve by the range of the abscissa. For example the A_z for the FROC curve in Figure 8.3 is computed by dividing the area under the curve by 2.5. This analysis is used in all results to follow.

8.4 Detection criteria

Evaluating the performance of a microcalcification CAD scheme requires certain criteria for determining a TP and FP cluster detection. Firstly, genuine clusters must be identified by an expert radiologist who provides a ground truth showing the location and size of each cluster. The most common criteria found in the literature for counting the number of true-positive detections is to regard a cluster as correctly detected if three or more microcalcifications associated with the cluster are found within the region marked as containing clusters by the expert radiologist [71, 115, 150]. All other clusters detected are considered to be false-positives (FP). For example, Figure 8.4(a) shows part of a mammogram image containing a single genuine microcalcification cluster. Overlayed on this image is the asso-

ciated ground truth as marked by an expert radiologist. Figure 8.4(b) shows the potential microcalcifications identified by the CAD scheme described in Chapter 6. Figure 8.4(c) shows two clusters detected by the new hybrid clustering algorithm detailed in Chapter 7. The first cluster (points marked with '×') is considered to be a TP detection since more than three individual microcalcifications associated with the cluster are inside the ground truth. The second cluster (points marked with '•') however is considered to be a false-positive detection.

8.5 Image databases

Images from two mammographic modalities have been used in developing the new microcalcification CAD scheme: digitised film-screen mammograms and direct digital mammograms. The image characteristics of these modalities differ significantly, and various stages of the CAD system have been developed separately for each imaging method. For this reason, when determining the accuracy of the CAD scheme, a performance evaluation is carried out on two separate databases, one consisting of digitised film-screen mammograms and the other consisting of direct digital mammograms.

For application to digitised film-screen mammograms, FSD-B is used (see Section 2.2). This database contains 50 mammogram images containing a total of 37 malignant microcalcification clusters (no benign microcalcification clusters). The images containing microcalcification clusters have an associated *overlay* image which provides annotations marking the location and size of each cluster. It was found that the annotations, which were provided by expert radiologists, were extremely generous in size (see Figures 7.14 to 7.16). For this reason a true-positive (TP) cluster is regarded as detected if all microcalcifications associated with the cluster are within the marked area on the overlay image. All other clusters found are considered to be false-positives (FP). The criteria for a TP detection is thus very restricting and ensures the clusters found have accurate localisation. No verification of the detection of individual microcalcifications is performed.

For application to direct digital mammogram images the direct digital database described in Section 2.2 is used. The data set consists of 124 mammograms from 33 patients. Twenty eight of the mammograms contain a single microcalcification cluster (14 benign and 14 malignant). Labelling of the microcalcification clusters was performed by an expert radiologist who provided both the pathology and location of each cluster. For this database a true-positive (TP) cluster is regarded as detected if three or more microcalcifications are found in the ground truth marked by the expert radiologist (see Section 8.4). All other clusters found are considered to be false-positives (FP). Once again no verification of the detection of individual microcalcifications is performed.

The CAD scheme described in this thesis is designed for the detection of microcalcification clusters. It does not provide a classification of the clusters into benign or malignant types. For this reason, in the following performance evaluation no distinction is made between benign or malignant microcalcification clusters. Both are considered to be true-positive detections. Also, due to the small size of the mammogram databases, separate testing and training data sets have not been used. In other words the same set of images (or at least a subset of them) were used while developing the CAD scheme as are employed here.

8.6 FROC implementation

Applying FROC analysis to a CAD scheme requires multiple operating points to be accurately determined. As mentioned previously detection algorithms have parameters that can be varied to alter the TP and FP rates. Each set of parameter values result in a different operating point. For the CAD scheme described in Chapter 6 the most obvious parameter to adjust is the microcalcification external field parameter ($A(2)$, described in Section 6.5). Setting $A(2)$ low increases the likelihood of a pixel being labelled as a microcalcification; conversely, setting $A(2)$ high decreases the number of pixels labelled as microcalcification. This provides a convenient adjustment in the sensitivity level for the microcalcification CAD scheme.

The FROC curve is obtained as follows:

1. Initialise parameter $A(2)$.
2. Apply the CAD scheme to all mammograms in the database being tested.
3. Record the number of TP, FN and FP clusters.
4. Use the recoded information to locate a single point on the FROC curve.
5. Adjust parameter $A(2)$ and repeat steps 2 to 4.

This iterative process is continued until sufficient operating points are found covering a large range of sensitivity levels. A smooth curve is then interpolated between the operating points, to produce the FROC curve. It should be noted that there are different methods for generating FROC curves than the method described above. For example, Anastasio *et al* developed a multi-objective genetic algorithm to produce the best FROC curve for a given dataset [4]. The FROC generation method used here however is much simpler and is considered to be sufficient for comparing the different CAD implementations.

Estimation of the area under the FROC curve was obtained using the trapezoidal rule. The trapezoidal rule systematically underestimates the area under each FROC curve. However, in the results to follow competing methods generate a similar number and distribution

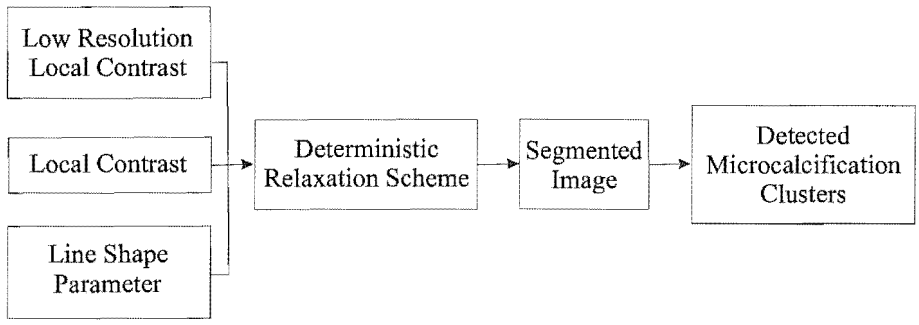


Figure 8.5 Implementation of CAD scheme.

of operating points, therefore the bias in the A_z value for each method is approximately the same and a comparison is valid. As mentioned in Section 8.3, A_z is computed by normalising the area under the FROC curve by the range of the abscissa. To compare A_z values for different FROC curves the area must be determined using the same abscissa range. In the results to follow the abscissa range is chosen to be $0.05 - 2$ false-positives per image. For a false-positive rate of less than 0.05/image the FROC curve is considered to be too steep for accurate evaluation, and the sensitivity too low for a practical CAD system. Additionally, a system with a false-positive rate greater than 2/image is not considered very useful.

8.7 Digitised film-screen evaluation

This section provides a performance evaluation of the microcalcification CAD scheme when applied to the mammograms in FSD-B. Implementation of the CAD scheme is achieved using the system shown in Figure 8.5. Three different features are used to represent the original mammogram data: the output of a line shape parameter and the local contrast at two different spatial resolutions [71]. Recall that calculation of the local contrast involves the development of an accurate high frequency noise equalisation technique (see Chapter 3). The line shape parameter which is designed to highlight CLS is described extensively in Chapter 5. These features are input to a deterministic relaxation scheme (see Chapter 6) which labels pixels as either background, microcalcification or CLS. Identification of the microcalcification pixels allows clusters to be found using the new hybrid clustering algorithm in Chapter 7.

The performance evaluation of the CAD scheme compares the various implementations of the high frequency noise equalisation technique and the line shape parameter, both of which are important in distinguishing microcalcifications from other structures within the

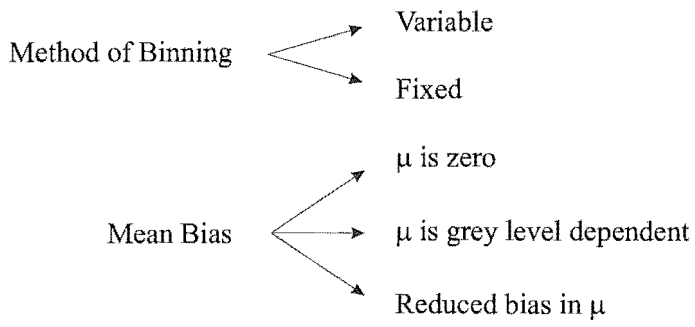


Figure 8.6 Implementation options for the digitised film-screen noise equalisation technique

breast. Starting with the noise equalisation technique, Figure 8.6 illustrates the different implementation options that were presented in Chapter 3.

It has been shown that the high frequency image noise varies as a function of the grey level in a mammogram [71,150]. To accurately measure the noise level over the entire pixel value range, the greyscale was divided into non-overlapping but adjacent bins for noise equalisation. Two different binning methods were investigated: fixed size binning and variable size binning. From the preliminary results shown in Chapter 3 it appeared that the variable size binning scheme provided a more accurate noise model compared to the fixed size bins, especially in high grey level regions of the mammogram. The most appropriate binning scheme for microcalcification detection has been determined by testing the performance of the CAD scheme using the two different binning techniques. The comparison is presented in subsection 8.7.1.

The second noise model variation investigated in Chapter 3 involved removing the mean bias in the histograms of local contrast within each bin. It was found that the histograms were not symmetric around zero, particularly if the number of pixels per bin was small. Three different methods of dealing with the bias in the local contrast histograms were investigated. The first was to assume the bias was negligible, while two other correction schemes attempted to reduce the bias. In the first correction scheme the mean bias was calculated as a function of the greyscale. The second scheme reduced the bias in the local contrast histograms by changing the sampling scheme prior to noise equalisation. The effect of reducing the bias on the detection of microcalcification clusters can best be determined by testing the microcalcification CAD scheme performance incorporating the three different techniques mentioned above. This is shown in subsection 8.7.2.

The final evaluation carried out in this section involves the two line shape parameters described in Chapter 5: the Hough shape parameter and the phase congruency shape

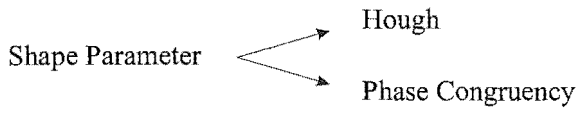


Figure 8.7 Implementation options for the line shape parameter.

parameter. Both these parameters were designed to distinguish CLS from microcalcifications (see Figure 8.7). Karssemeijer developed the Hough shape parameter using the local probability density function of gradient directions to determine the presence or absence of local line-like features. The phase congruency shape parameter on the other hand utilised and extended a method proposed by Kovesi [86], implementing feature detection based on local phase information. The preliminary results shown in Chapter 5 indicate that the phase congruency shape parameter provides a better distinction between CLS and microcalcification. This would suggest that the phase congruency method should give superior microcalcification detection performance. The effect of the shape parameter on the detection of microcalcification clusters can best be determined by testing the microcalcification CAD scheme performance incorporating the two different shape parameter techniques mentioned above. This is shown in subsection 8.7.3.

8.7.1 Method of binning

Comparing the CAD performance based solely on the binning methods described above requires all other implementation options to be fixed. For the purposes of this comparison the mean bias is calculated as a function of the greyscale and the phase congruency shape parameter is used. Figure 8.8(a) illustrates the two binning options to be tested as well as the fixed options chosen for the mean bias and shape parameter.

Using FROC analysis, two FROC curves were obtained and are shown in Figure 8.8(b). The top curve corresponds to the CAD performance when using a variable size binning scheme while the lower curve corresponds to the fixed size bins. These results indicate that variable size binning for noise equalisation provides superior microcalcification cluster detection performance. For example, at a true-positive rate of 0.91, the number of false-positives per image decreases from 1.1 to 0.8. To provide more quantitative results the A_z values associated with each curve were calculated. For an abscissa range of $0.05-2$ false-positives per image, $A_z = 0.84$ for the fixed size binning scheme, and $A_z = 0.89$ for the variable bin size.

The performance analysis shown here supports the preliminary results in Chapter 3 where the variable size binning method allowed more reliable estimation of the noise level in the

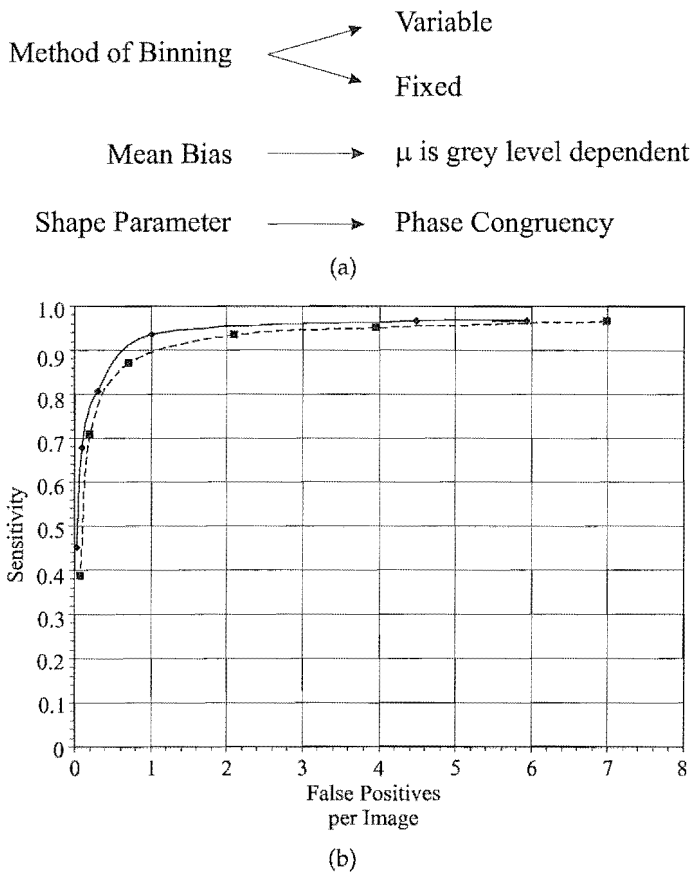


Figure 8.8 (a) Implementation options for comparing the noise equalisation binning scheme. (b) FROC plot showing the CAD performance using different methods of binning, (♦) corresponds to the variable size binning scheme and (■) corresponds to the fixed size binning scheme.

high intensity regions of a mammogram. This has lead to better microcalcification detection performance hence reinforcing the fact that an accurate high frequency noise model is important in the detection of microcalcifications. Veldkamp *et al* also carried out a study on various methods of greyscale binning concluding that the variable size binning scheme as described in Chapter 3 provides superior microcalcification detection performance [150]. The results presented here support Veldkamp *et al*'s conclusions.

8.7.2 Mean bias

To compare the CAD performance based on various methods of reducing the mean bias requires all other implementation options to be fixed. In the previous subsection it was found that variable size binning for noise equalisation provides superior microcalcification detection performance compared to fixed size binning. Therefore, for the purposes of this comparison the variable size binning scheme is used along with the phase congruency shape parameter. Figure 8.9(a) illustrates the three mean bias options to be tested as well as the fixed options chosen for the binning scheme and shape parameter.

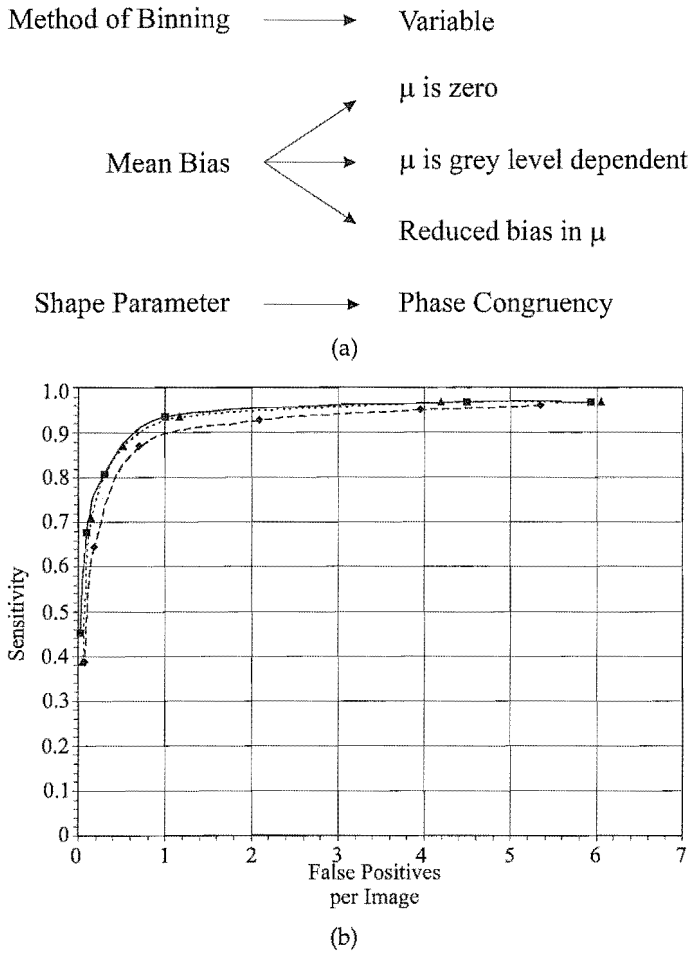


Figure 8.9 (a) Implementation options for comparing the reduction in mean bias. (b) FROC plot showing the CAD performance using different methods of reducing the mean bias, (♦) illustrates when the mean bias is assumed to be zero, (▲) shows when the mean bias is reduced by changing the sampling scheme and (■) shows when the mean bias is calculated as a function of the greyscale.

Using FROC analysis, three FROC curves were obtained and are shown in Figure 8.9(b). The lowest curve represents the CAD systems performance when the mean bias is assumed negligible. The top two curves correspond to the two correction schemes. For an abscissa range of $0.05 - 2$ false-positives per image, $A_z = 0.83$ when the mean bias is assumed negligible, $A_z = 0.89$ when the bias is calculated as a function of the greyscale and $A_z = 0.88$ when the mean bias is reduced by changing the sampling scheme prior to noise equalisation. This provides clear evidence that compensating for the mean bias in noise equalisation improves microcalcification detection performance. It should also be noted that while the two correction methods show very similar detection performance, slightly superior results are obtained when calculating the mean bias as a function of the greyscale.

Veldkamp *et al* also carried out a study on microcalcification CAD performance using the above mentioned methods of compensating for the bias. The author concluded that the two

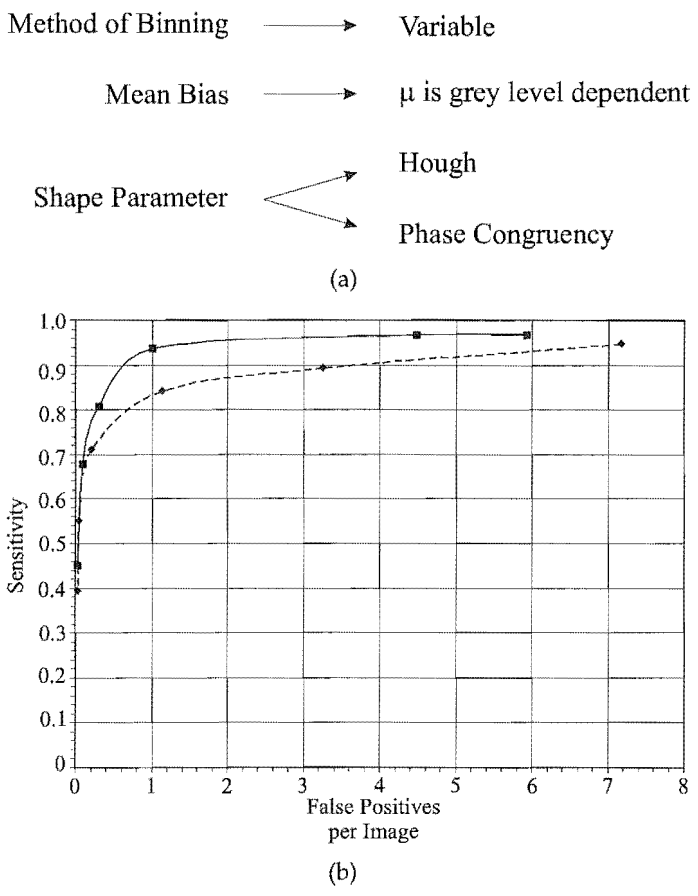


Figure 8.10 (a) Implementation options for comparing the line shape parameter. (b) FROC plot showing the CAD performance using different line shape parameters, (♦) corresponds to the Hough shape parameter and (■) corresponds to the phase congruency shape parameter.

compensation methods for reducing the bias provide superior detection performance [150]. Once again the results presented here support Veldkamp *et al*'s conclusions.

8.7.3 CLS shape parameters

Comparing the CAD performance based solely on the line shape parameter requires all other implementation options to be fixed. For the purposes of this comparison the mean bias is calculated as a function of the greyscale and the variable size binning scheme is used. These options were chosen as they provided the best CAD performance in the previous subsections. Figure 8.10(a) illustrates the two shape parameter options to be tested as well as the fixed options chosen for the mean bias and the binning scheme.

Using FROC analysis, two FROC curves were obtained and are shown in Figure 8.10(b). The top curve corresponds to the CAD performance when the phase congruency shape parameter is used while the lower curve corresponds to the Hough shape parameter. These results indicate the phase congruency shape parameter provides superior microcalcifica-

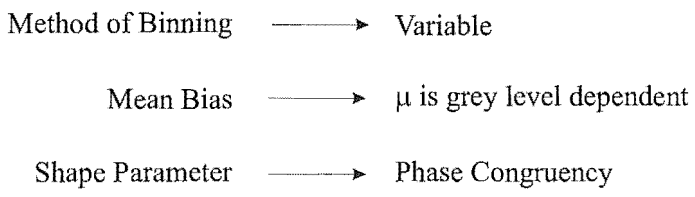


Figure 8.11 Final implementation options for the CAD scheme applied to digitised film-screen mammograms.

tion cluster detection performance. For example, at a true-positive rate of 0.91, the number of false-positives per image decreases from 4 to 0.8. For an abscissa range of $0.05 - 2$ false-positives per image, $A_z = 0.80$ for the Hough shape parameter, and $A_z = 0.89$ for the phase congruency shape parameter. These values indicate a marked improvement in the microcalcification CAD results when using the phase congruence method.

The performance analysis shown here supports the preliminary results in Chapter 5 where the phase congruency shape parameter is seen to provide a better distinction between CLS and microcalcification. It should also be noted that the variation in CAD performance for the two shape parameters is significantly greater than the performance variation associated with the noise equalisation technique. This implies that determination of an accurate shape parameter capable of distinguishing CLS from microcalcifications is an important step in any microcalcification CAD scheme.

To summarise the results presented in this section, Figure 8.11 shows the final implementation options providing the most accurate CAD system performance when applied to digitised film-screen mammograms.

8.8 Direct digital evaluation

This section provides a performance evaluation of the microcalcification CAD scheme when applied to mammograms in the direct digital database. Implementation of the CAD scheme is achieved using the same system shown in Figure 8.5. For the purposes of this evaluation the phase congruency shape parameter is used as it was found to provide superior microcalcification detection performance for the digitised film-screen modality.

Due to the significantly different image characteristics of the digitised film-screen and direct digital modalities different high frequency noise equalisation techniques have been employed. Chapter 4 describes the noise equalisation technique developed specifically for the direct digital modality. The performance evaluation of the CAD scheme to follow, focusses on various implementations of the high frequency noise equalisation technique.

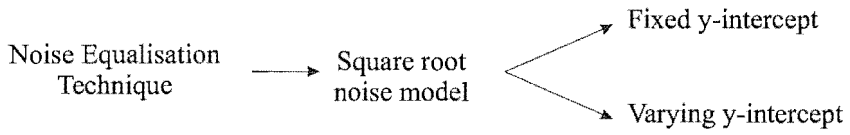


Figure 8.12 Implementation options for the direct digital noise equalisation technique.

Figure 8.12 illustrates the different implementation options that were investigated in Chapter 4.

The assumed quantum noise dominance of the direct digital system and the linear characteristic curve that this modality exhibits ensures that a simple square root noise model provides accurate estimation of the noise with respect to grey level. Estimation of the square root noise model requires two parameter values to be determined. One of these parameters corresponds to the noise present at zero exposure (zero grey level). In a strictly quantum limited system the noise present at zero exposure (zero grey level) would be zero. However, it was found in Chapter 4 that the noise at zero exposure had a very small positive value possibly due to a small additional noise source present within the direct digital imaging system. In the next subsection the CAD performance based on two square root noise model implementations are investigated. First, the square root model is reduced to finding a single parameter by fixing the noise value at zero exposure (fixed y-intercept). Second, both square root noise model parameters are estimated (varying y-intercept) which should theoretically provide a more accurate noise model for the equalisation technique.

8.8.1 Noise equalisation methods

The performance evaluation presented here provides a comparison between the square root noise model implementations described above and the noise model developed specifically for digitised film-screen mammograms (see Chapter 3). For the purposes of this comparison the implementation options shown in Figure 8.11 are used for the digitised film-screen noise equalisation scheme. Additionally, these two noise equalisation techniques are compared with the results obtained when noise equalisation is omitted. Figure 8.13(a) illustrates the four noise equalisation options to be investigated.

The FROC curves shown in Figure 8.13(b) evaluate the detection performance of the CAD system utilising different noise equalisation schemes. Note that the abscissa range in Figure 8.13(b) is much smaller than previous FROC curves presented in this chapter. The performance data collected for the comparison shown here ensures an accurate evaluation of the CAD scheme at low false-positive detection rates, which is the most significant region of any FROC plot (see Section 8.3).

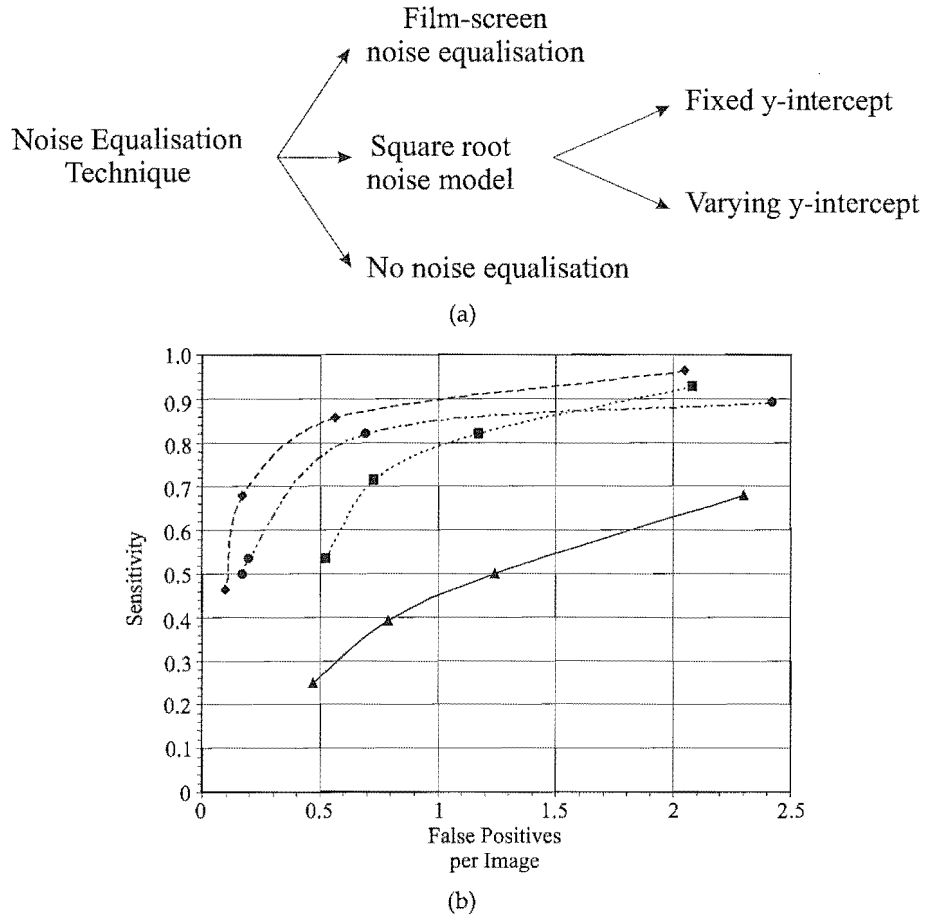


Figure 8.13 (a) Implementation options for comparing the noise equalisation technique. (b) FROC curves showing the CAD performance using different noise equalisation techniques, (♦) shows the square root noise model with a varying y-intercept, (•) shows the square root noise model with a fixed y-intercept, (■) shows the noise equalisation technique developed for digitised film-screen mammograms and (▲) is when noise compensation is omitted.

The lowest curve in Figure 8.13(b) represents the results obtained when noise equalisation is omitted. This provides clear evidence that high frequency noise compensation is important for the detection of microcalcifications. An improvement is shown in detection performance when using the noise equalisation scheme developed in Chapter 3, however the best performance is obtained when implementing the square root noise model. The CAD performance for both a fixed and varying y-intercept for the square root noise model provides the top two FROC curves in Figure 8.13(b). These two curves indicate that fixing the y-intercept does not greatly reduce the sensitivity of the CAD system. When comparing the square root noise model with the film-screen noise equalisation scheme, at a true-positive fraction of 0.91 the number of false-positives per image decreases from 1.8 to 1.2.

To provide more quantitative results the A_z values associated with each curve have been calculated. For an abscissa range of $0.05-2$ false-positives per image, $A_z = 0.41$ when noise equalisation is omitted, and $A_z = 0.68$ for the digitised film-screen noise equalisation technique. For the square root noise model with a fixed y-intercept $A_z = 0.77$ and $A_z = 0.85$ for the varying y-intercept. These results prove that the square root model based approach which direct digital allows leads to a more robust estimation of the high frequency image noise and thereby better microcalcification detection.

8.9 Discussion

It has been shown that accurate detection of microcalcification clusters is highly dependent on correct high frequency noise equalisation and the determination of an accurate line shape parameter. The performance evaluation carried out in this chapter has identified the most appropriate CAD implementations for both digitised film-screen and direct digital mammograms.

FROC analysis was used in determining the CAD systems performance because of its ability to accommodate multiple abnormalities per image and its requirement for correct localisation of microcalcification clusters. Unfortunately, no validated methods are currently available for the statistical testing of FROC curve estimates. Some techniques for FROC analysis are described in the literature, however many of these approaches require several strong assumptions concerning the statistical properties of the TP and FP detections that are not likely to be met in practice. For this reason the FROC method has not been widely employed [21, 22, 109]. Calculation of the A_z index is one of the only measures believed to be suitable for FROC evaluation performance. Care must be taken in determining this value to ensure comparisons can be made between competing methods.

Although a direct comparison was not made between the CAD performance for the digitised film-screen modality and the CAD performance for the direct digital modality, observation of the FROC curves show that using the noise equalisation technique developed specifically for digitised film-screen mammograms (see Chapter 3) produces a higher CAD performance for the digitised film-screen database compared to the direct digital database. This is to be expected due to the significantly different image characteristics of the two modalities. The results indicate that it is important to develop a noise model specific to each imaging method in order to provide optimal microcalcification detection performance.

Comparing the CAD scheme developed here with previously developed microcalcification detection schemes shows that the performance of this method is highly competitive. As mentioned in Section 6.2 the best CAD results reported to date have a sensitivity of $85-90\%$

and a false-positive detection rate of approximately 1 false-positive per image. The best results presented in this chapter offer a slight improvement giving a sensitivity of 91% at a false-positive detection rate of 0.8 false-positives per image. It should be noted that an accurate comparison between microcalcification CAD schemes however is unavailable due to the lack of a common database or common objective criteria.

8.10 Summary

An extensive performance evaluation of the CAD scheme developed in previous chapters has been carried out on two separate databases, one consisting of digitised film-screen mammograms and the other consisting of direct digital mammograms. An understanding of the evaluation criteria used for CAD schemes helps determine the most appropriate performance analysis for this application, the free-response operating characteristic (FROC) analysis. The aim was to find the best system performance based on various algorithm implementations detailed in previous chapters.

For the digitised film-screen modality two variants were investigated for dividing the greyscale into bins. It appears that the detection results related to noise estimation in fixed size bins can be improved by using bins consisting of a variable number of pixels. The variable binning scheme allows more reliable estimation of the noise level in the high intensity regions of a mammogram, which leads to improved microcalcification detection results.

A number of approaches were also investigated to remove the mean bias in the histograms of local contrast within each bin. Compensation for the bias during the noise equalisation process improves microcalcification detection. Specifically, assuming the mean bias to be grey level dependent gave the best CAD results.

The final implementation option investigated for the digitised film-screen modality was the line shape parameter. Two methods were compared: the Hough shape parameter and the phase congruency shape parameter. The phase congruency method provides superior microcalcification detection performance, supporting preliminary results that show a better distinction between CLS and microcalcification when using this method.

For the direct digital modality the performance evaluation gave a comparison between the square root noise model and the noise model developed specifically for digitised film-screen mammograms. Also shown was a comparison with the results obtained when noise equalisation was omitted. Noise equalisation, performed with the use of the square root model, produces more accurate detection of microcalcifications in direct digital mammograms than either no noise equalisation or the more complex method developed for digitised film-screen mammograms.

Chapter 9

Conclusions and future work

This thesis describes the development of a computer aided detection (CAD) scheme for the detection of microcalcification clusters in digital mammogram images. The research undertaken during the course of this PhD has led to a number of interesting results. They are discussed in detail in the relevant chapters and are summarised here. In addition, suggestions for future research are presented.

9.1 Conclusions

Detecting breast cancer as early as possible increases the survival chance of patients and consequently breast cancer screening programs have been introduced in a number of countries around the world. One of the most important radiologic tasks in the screening process is the detection of microcalcification clusters. There is evidence that radiologists perform better at detecting subtle microcalcifications when provided with cues from a CAD scheme marking suspicious regions. Moreover, the display of a computer detected microcalcification pattern is found to be useful in the classification of microcalcifications. The CAD scheme detailed in this thesis is designed to highlight the location of both benign and malignant microcalcification clusters potentially assisting radiologists in improving the effectiveness of mammography screening.

There are essentially two imaging modalities available in digital mammography: digitised film-screen mammograms and direct digital mammograms. The modalities possess considerably different image characteristics which affect, for example, the image contrast and required patient dose. A film-screen system has a non-linear response to x-ray exposure, which is a significant limitation of the modality, while a direct digital system has an approximately linear response to x-ray photon flux, regardless of the level of exposure. Because of the difference in image characteristics certain stages of the CAD scheme are developed sep-

arately for each modality, thus ensuring optimal microcalcification detection performance for each imaging method.

Extraction of quantitative features for the detection of microcalcification clusters requires accurate and robust segmentation methods for separating the breast tissue from the pectoral muscle and other image regions outside the breast (background). Removing background pixels in a mammogram significantly reduces the processing time and improves the efficiency of subsequent computer programming stages. Automatic determination of the pectoral muscle is useful in restricting the search space for cancer detection, as well as determining mammogram adequacy and mammogram-pair registration. Chapter 2 described the development of accurate and robust segmentation methods for both the skin-air boundary and the pectoral muscle boundary. A multi-stage algorithm provides detection of the breast boundary in film-screen mammograms, while the image properties of the direct digital modality allow a simpler global thresholding technique to be applied in locating the breast tissue/air boundary. For detection and segmentation of the pectoral muscle boundary, a straight line Hough transform is found to be sufficient for both modalities.

Accurate detection of individual microcalcifications is limited by the high spatial frequency noise present in digital mammogram images. In order to accurately detect microcalcifications, measurement of the high frequency image noise is of crucial importance. A noise model for the digitised film-screen modality was developed in Chapter 3. The noise components present in digitised film-screen mammograms identify the need for a complex noise model that is grey level dependent. Based on a method developed by Veldkamp *et al*, an accurate adaptive approach, referred to as noise equalisation, was taken which removes the dependency of the noise on the grey level.

Optimising the noise equalisation technique (also detailed in Chapter 3) was carried out by investigating a number of alternative implementations aimed at improving the reliability of the noise estimates. First, a method of truncating the local contrast probability density function prior to estimation of the noise statistics was considered. Truncation ensures that outliers associated with feature boundary pixels do not affect the noise estimation and provides a more reliable and accurate representation of the noise level. Second, two methods for dividing the greyscale into bins were investigated, one using a fixed number of pixels per bin and the other using a variable number of pixels. The variable size binning scheme provides a more accurate noise model compared to the fixed size bins, especially in high grey level regions of the mammogram. Finally, three methods of compensating for the bias in the local contrast distributions were investigated. The effect of bias compensation on the accuracy of the noise model could not be directly determined due to the complexity of the noise in this modality. However, it was found that improved microcalcification detection performance is achieved when compensating for the bias in the noise equalisation scheme.

In contrast to digitised film-screen mammograms, direct digital mammograms do not suffer from a complex highly variable relationship between intensity and noise variance. A near linear relationship exists between pixel value and exposure for the direct digital modality. This allows a simple noise model to be applied to direct digital mammograms, leading to more robust estimation of the image noise. In Chapter 4 a noise model was developed for direct digital mammograms based on the noise equalisation technique described in Chapter 3. In the model it is assumed that quantum noise is the dominant noise source in this modality. The combination of quantum noise dominance and the linear detector response of the direct digital system provides a simple square root relationship between the noise and the grey level.

Experimental results show that the square root noise model is accurate for both direct digital step wedge images and direct digital mammograms. Truncating the local contrast distributions prior to noise equalisation provides a more reliable and accurate representation of the noise as also found for the digitised film-screen modality. Best results for the noise model are obtained if the pixels associated with the uncompressed area of the breast are excluded from the estimation. There is significant deviation from the square root model in these pixels, suggesting that noise processes other than quantum noise may be playing a role. Fixing the residual contrast standard deviation at zero exposure simplifies the noise estimation process to finding a single parameter, although this causes a slight decrease in noise model accuracy. Overall, the small error values reported in Chapter 4 confirm that the square root noise model is accurate for application to direct digital images. This is believed to be the first report discussing the detection of microcalcifications in direct digital mammogram images that takes full advantage of the linearity of the system.

The local operations used for the detection of microcalcifications often respond strongly to curvilinear structures (CLS). When detecting microcalcification clusters, it is important that the appearance of both microcalcifications and CLS are modelled. Chapter 5 detailed the development of a shape parameter designed specifically to highlight CLS while simultaneously suppressing microcalcifications. An existing method proposed by Karssemeijer uses an algorithm resembling the Hough transform, though its limitations include high sensitivity to the parameters chosen, lack of adaptability to features of varying size, and high execution time. The inherent properties of CLS such as variable size and orientation suggest the need for a multiresolution shape parameter. This was provided in Section 5.4 with the development of a new shape parameter based on the method of phase congruency. The method treats features in terms of their Fourier components and uses the local property measures from a multiresolution wavelet filter bank to develop a shape parameter highlighting CLS and suppressing microcalcifications. The results indicate that the phase congruency shape parameter is superior because of its adaptability to the varying size and orientation of CLS and lower execution time.

In Chapter 6 the overall microcalcification CAD system was developed, which extends a well known method proposed by Karssemeijer. In the scheme a stochastic model enables the classification of individual pixels within a mammogram into separate classes based on Bayesian decision theory. A principal advantage of the approach is that all the information available, i.e. the image data, the current labelling and prior beliefs, is exploited simultaneously. The inclusion of a line shape parameter lead to an efficient way of distinguishing curvilinear structures from faint microcalcifications. This distinction is improved with the introduction of a connectivity term. The connectivity utilises the fact that pixels classified as CLS are likely to occur along connected lines within the image, thus improving microcalcification detection by reducing the number of individual false positive detections related to CLS.

The computer aided detection scheme described in Chapter 6 produces a binary image that identifies potential individual microcalcification objects. However, since individual (isolated) microcalcifications are in most cases not clinically significant, it is important that a computer detection scheme distinguish between clustered and individual microcalcifications. The binary images resulting from the CAD scheme were used as input to a clustering algorithm detailed in Chapter 7. The scheme aims to detect true microcalcification clusters while ignoring isolated individual microcalcifications and falsely identified signals.

The clustering process involves the combination of a mode clustering algorithm formerly proposed in the literature and a refinement algorithm based on Voronoi sets. Mode clustering with conservative choices for the controlling parameters is an effective and efficient method for the first stage of the hybrid scheme, although it suffers from a high dependency on the parameters chosen. For this reason, a second stage was applied that employs Voronoi neighbours to refine the clusters already identified. The results obtained by the new hybrid clustering scheme are encouraging, with the majority of genuine microcalcification clusters correctly detected. A few false positive detections are made but in most of these cases the layout of the points is consistent with the expected properties of a genuine cluster. It should be noted that the clusters determined in this process are not classified into benign or malignant cluster types.

To determine the accuracy of the CAD system and hence the potential contribution to the mammography screening process, a performance evaluation was conducted and the results presented in Chapter 8. The aim was to determine the best system performance based on various algorithm implementations from Chapters 2 to 7. Free-response receiver operating characteristic (FROC) analysis was used to determine the CAD system's performance because of its ability to accommodate multiple abnormalities per image and its requirement for correct localisation of microcalcification clusters. Due to the significantly different image characteristics of the digitised film-screen and direct digital modalities, two separate

mammogram databases were used, one consisting of digitised film-screen mammograms and the other consisting of direct digital mammograms.

The results presented in Chapter 8 strongly reinforce the findings of Chapters 2 to 7 for both the noise equalisation techniques and shape parameter methods. For the digitised film-screen modality, the detection results related to noise estimation in fixed size bins are improved by using bins consisting of a variable number of pixels. Additionally, compensation for the bias during the noise equalisation process improves microcalcification detection. Finally, the phase congruency shape parameter provides superior detection performance over the Hough shape parameter.

For the direct digital modality, the performance evaluation gives a comparison between the square root noise model and the noise model developed specifically for digitised film-screen mammograms, as well as the results obtained when noise equalisation is omitted. Noise equalisation, performed with the use of the square root model, produces more accurate detection of microcalcifications in direct digital mammograms than either no noise equalisation or the more complex method developed for digitised film-screen mammograms. These results indicate that it is important to develop a noise model specific to each imaging method in order to provide optimal microcalcification detection performance.

Comparing the CAD scheme developed here with previously developed microcalcification detection schemes shows that the performance of this method is highly competitive. The best CAD results reported to date have a sensitivity of $85 - 90\%$ and a false positive detection rate of approximately 1 false positive per image. The best results presented in this thesis offer a slight improvement giving a sensitivity of 91% at a false positive detection rate of 0.8 false positives per image. It should be noted that an accurate comparison between microcalcification CAD schemes however was unavailable due to the lack of a common database or common objective criteria.

9.2 Future work

The data set of mammogram images used in the development and evaluation of this CAD scheme provides only a small sample of the possible spectrum of microcalcification clusters in a screening population. A more in depth analysis of the scheme using a larger image database would determine the robustness of the system and provide a more realistic indication of the scheme's potential contribution to the mammography screening process. On a related topic, testing the microcalcification detection performance of an expert radiologist, when using cues from the CAD scheme marking suspicious regions, would also indicate the CAD scheme's potential contribution to breast screening. If the scheme is capable of detecting subtle microcalcification clusters that a radiologist would otherwise overlook, the efficacy of screening would improve.

Statistical testing of the algorithms developed in this thesis is limited due to the lack of a known model for comparison with the research carried out. For example, no ground truth is available for the breast tissue or pectoral muscle areas, therefore accuracy of the outlines require visual inspection. Additionally, the unknown noise model for the digitised film-screen modality means the accuracy of the model described in Chapter 3 can not be determined. One aspect of the thesis that would greatly benefit from rigourous statistical testing is the FROC analysis in Chapter 8. The small size of the image databases limit the analysis that can be carried out. Statistical comparison of different CAD implementations require algorithms such as jackknifing and bootstrapping to be applied in the context of database selection, testing and training [32, 38]. These algorithms allow the database to be divided into a number of subsets n. The CAD scheme is trained n times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the FROC curve. To determine the statistical significance between a set of curves each associated with different algorithm implementations, statistical testing methods such as the paired t-test can be used. It seems likely that a significantly larger database than those used in this thesis would be required.

In almost all digitised film-screen mammograms, a band of pixels with a higher signal intensity, up to $2cm$ in width, is found along the posterior edge of the direct-exposure area. In some instances this artifact is so severe that it completely masks the adjacent breast border, thus causing a decrease in breast boundary detection accuracy in this region (as was found in Chapter 2). Development of an algorithm to remove these pixels from the segmented breast tissue area would result in a decrease in processing time for subsequent computer programming stages. Execution of the algorithm however, must be simple enough to ensure the increase in computational load is not deemed too great.

One limitation of the noise model developed in Chapter 4 for direct digital mammograms, is the significant deviation from the model in the peripheral area of the breast, where the tissue is uncompressed. It is argued that in the peripheral area of the breast the image is not photon limited and noise processes other than quantum noise may play a significant part. Furthermore, recent experiments performed by researchers in Nijmegen show that irregularity of the skin surface, amplified near the breast edge because of the small angle between the skin surface and the incident x-ray beam also contributes significantly to the deviation from the noise model. More work is needed to fully explain the observations and determine the additional noise present in this area. This could potentially lead to an improved microcalcification detection performance at the breast edge.

As mentioned in Chapter 5 some of the inherent properties of both microcalcifications and CLS make it difficult for a line-like shape parameter to distinguish between these two features. For example, within a microcalcification cluster individual calcifications are some-

times positioned very close to one another and may even overlap, causing a shape parameter to misinterpret the pixels as belonging to small CLS. The same misinterpretation can occur with slightly elongated microcalcifications which commonly occur in mammography. It is unclear how to overcome these inherent problems and more work must be carried out to improve the distinction between the two features types, potentially improving microcalcification detection performance.

The efficiency of the computer aided detection scheme has not been addressed in this thesis. Obviously the rate at which a CAD scheme evaluates a mammogram is important for consideration in a screening process. On average a radiologist screens approximately 100 cases per hour and up to a maximum of 300 cases per day. While a CAD scheme does not need to operate at this high rate, due to the ability of the scheme to work off-line, it still needs to process images in less than $2 - 3$ minutes to be an effective screening tool. The CAD scheme developed here was implemented in MATLAB which is a relatively inefficient programming language. Future work would require the optimisation of this scheme for application to a screening program.

Another application for CAD in mammography includes the detection of stellate lesions. These lesions are identified by their irregular borders and thin radiating pattern of spicules that emanate from the body of the lesion. The spicules associated with stellate lesions have properties similar to that of CLS. Thus, application of the shape parameter developed in Chapter 5 could provide a means of detecting these spicules. Additionally, characterising breast tissue parenchymal patterns using computer algorithms enables different breast patterns to be determined. This is associated with differing levels of risk for breast cancer developing. Here is another area where the detection of locally linear features would be important.

As yet, the method of noise equalisation presented in Chapters 3 and 4 is not seen to have application to any other medical imaging devices. However, as the utilisation of flat panel solid-state detectors continues to increase, other applications may arise where relatively high frequency signals are to be detected in quantum noise.

Appendix A

Digital Database for Screening Mammography (DDSM)

All images obtained from this database were downloaded from an anonymous ftp site at [58]: *ftp://figment.csee.usf.edu/pub/DDSM/cases/.*

A.1 Film-Screen Dataset A (FSD-A)

Table A.1 DDSM images used to develop the breast-air and pectoral muscle boundary segmentation algorithms.

Folder	Case	Image
Cancer01	0001	RC,RM
Cancer01	0002	LC,LM
Cancer01	0004	RC,RM
Cancer01	0006	LC,LM,RC,RM
Cancer01	0007	LC,LM,RC,RM
Cancer01	0011	RC,RM
Cancer01	0012	RC,RM
Cancer01	0015	RC,RM
Cancer01	0016	RC,RM
Cancer02	0018	LC,LM
Cancer02	0022	RM
Cancer02	0023	LC,LM

continued on next page

continued from previous page

Folder	Case	Image
Cancer02	0028	LC,LM
Cancer02	0035	RC,RM
Cancer02	0038	LC,LM
Cancer02	0042	LC,LM
Cancer02	0052	RC
Cancer02	0057	RC,RM
Cancer02	0061	RC,RM
Cancer02	0063	RC,RM
Cancer02	0076	RC,RM
Cancer02	0091	LC
Cancer02	0092	RM
Cancer03	1000	LC,LM
Cancer14	1178	LC,LM

A.2 Film-Screen Dataset B (FSD-B)

Table A.2 DDSM images selected for development of the microcalcification CAD scheme.

Folder	Case	Image
Cancer01	0012	RC,RM
Cancer01	3082	LC,LM
Cancer02	0036	RC,RM
Cancer02	0045	RC,RM
Cancer02	0057	RC,RM
Cancer02	0063	RC,RM
Cancer02	0087	LC,LM
Cancer02	0090	LC,LM
Cancer02	0093	LC,LM
Cancer03	1002	RC
Cancer03	1022	LC,LM
Cancer03	1023	RM
Cancer03	1030	LC,LM
Cancer04	1001	LC,LM
Cancer04	1096	LC,LM

continued on next page

continued from previous page

Folder	Case	Image
Cancer04	1097	LC,LM
Cancer07	1213	LC,LM
Normal01	0002	LC,LM
Normal01	0003	LC,LM
Normal01	0005	LM
Normal01	0006	RC
Normal01	0009	RM
Normal01	0010	LC
Normal01	0011	LC
Normal01	0012	RC,RM
Normal01	0013	RC
Normal01	0014	LM
Normal01	0015	LM
Normal01	0016	LC
Normal01	0018	RM
Normal01	0020	LC
Normal01	0022	RM

Appendix B

Doubly truncated Gaussian

Using the same notation as that described in subsection 3.5.1 the population mean $E[C(k)]$ and variance $Var[C(k)]$ of a truncated Gaussian can be defined as [68]

$$E[C(k)] = \int_{-\infty}^{\infty} c\, p_T(c|k)\, dc \tag{B.1}$$

and

$$\text{Var}[C(k)] = \int_{-\infty}^{\infty} (c - E[C(k)])^2\, p_T(c|k)\, dc, \tag{B.2}$$

where $p_T(c|k)$, written in terms of the unit normal density, is

$$p_T(c|k) = \int_{c_{\text{low}}(k)}^{c_{\text{high}}(k)} \frac{1}{\sigma L} Z\left(\frac{c-\mu}{\sigma}\right) dc. \tag{B.3}$$

The parameter L can be calculated as

$$\begin{aligned} L &= \int_{c_{\text{low}}(k)}^{c_{\text{high}}(k)} p(c|k)\, dc \\ &= \Phi\left(\frac{c_{\text{high}}(k)-\mu}{\sigma}\right) - \Phi\left(\frac{c_{\text{low}}(k)-\mu}{\sigma}\right). \end{aligned} \tag{B.4}$$

In the following derivation both $E[C(k)]$ and $\text{Var}[C(k)]$ are related to the underlying Gaussian population statistics μ and σ. Beginning with the derivation of the mean $E[C(k)]$, substitute Equation B.3 into Equation B.1 to give

$$E[C(k)] = \int_{c_{\text{low}}(k)}^{c_{\text{high}}(k)} \frac{c}{\sigma L} Z\left(\frac{c-\mu}{\sigma}\right) dc. \tag{B.5}$$

At this stage a change in parameter is appropriate, thus let

$$x = \frac{c - \mu}{\sigma}, \tag{B.6}$$

which results in

$$E[C(k)] = \int_{\frac{c_{\text{low}}(k)-\mu}{\sigma}}^{\frac{c_{\text{high}}(k)-\mu}{\sigma}} \frac{\sigma x + \mu}{\sigma L} Z(x)\ \sigma\ dx$$

$$= \frac{\sigma}{L}\left[Z\left(\frac{c_{\text{low}}(k) - \mu}{\sigma}\right) - Z\left(\frac{c_{\text{high}}(k) - \mu}{\sigma}\right)\right] + \tag{B.7}$$

$$\frac{\mu}{L}\left[\Phi\left(\frac{c_{\text{high}}(k) - \mu}{\sigma}\right) - \Phi\left(\frac{c_{\text{low}}(k) - \mu}{\sigma}\right)\right].$$

Substituting Equation B.4 into Equation B.7 gives

$$E[C(k)] = \mu + \frac{Z\left(\frac{c_{\text{low}}(k)-\mu}{\sigma}\right) - Z\left(\frac{c_{\text{high}}(k)-\mu}{\sigma}\right)}{\Phi\left(\frac{c_{\text{high}}(k)-\mu}{\sigma}\right) - \Phi\left(\frac{c_{\text{low}}(k)-\mu}{\sigma}\right)}\ \sigma. \tag{B.8}$$

providing an equation relating the population mean and standard deviation of the underlying Gaussian with the population mean of the truncated Gaussian [68].

For the derivation of the variance $\text{Var}[C(k)]$, substitute Equation B.3 into Equation B.2 to give

$$\text{Var}[C(k)] = \int_{c_{\text{low}}(k)}^{c_{\text{high}}(k)} \frac{(c - E[C(k)])^2}{\sigma L} Z\left(\frac{c - \mu}{\sigma}\right) dc. \tag{B.9}$$

Using the same change in parameter shown in Equation B.6, $\text{Var}[C(k)]$ can be expressed as

$$\text{Var}[C(k)] = \int_{\frac{c_{\text{low}}(k)-\mu}{\sigma}}^{\frac{c_{\text{high}}(k)-\mu}{\sigma}} \frac{\left[(\sigma x - \mu) - E[C(k)]\right]^2}{\sigma L} Z(x)\ \sigma\ dx. \tag{B.10}$$

Expanding Equation B.10 into three terms results in

$$\text{Var}[C(k)] = \frac{\sigma^2}{L}\int_{\frac{c_{\text{low}}(k)-\mu}{\sigma}}^{\frac{c_{\text{high}}(k)-\mu}{\sigma}} x^2 Z(x)dx \quad + \frac{2(\sigma\mu - E[C(k)]\sigma)}{L}\int_{\frac{c_{\text{low}}(k)-\mu}{\sigma}}^{\frac{c_{\text{high}}(k)-\mu}{\sigma}} xZ(x)dx$$

$$+ \frac{(\mu^2 - 2E[C(k)]\mu + E[C(k)]^2)}{L}\int_{\frac{c_{\text{low}}(k)-\mu}{\sigma}}^{\frac{c_{\text{high}}(k)-\mu}{\sigma}} Z(x)dx \tag{B.11}$$

Each integral can be solved separately, providing

$$\mathrm{Var}[C(k)] = \Bigg[1 + \frac{((c_{\mathrm{low}} - \mu)/\sigma)Z((c_{\mathrm{low}} - \mu)/\sigma) - ((c_{\mathrm{high}} - \mu)/\sigma)Z((c_{\mathrm{high}} - \mu)/\sigma)}{\Phi((c_{\mathrm{high}} - \mu)/\sigma) - \Phi((c_{\mathrm{low}} - \mu)/\sigma)}$$

$$- \Big(\frac{Z((c_{\mathrm{low}} - \mu)/\sigma) - Z((c_{\mathrm{high}} - \mu)/\sigma)}{\Phi((c_{\mathrm{high}} - \mu)/\sigma) - \Phi((c_{\mathrm{low}} - \mu)/\sigma)}\Big)^2\Bigg]\sigma^2, \quad \text{(B.12)}$$

relating the population mean and standard deviation of the underlying Gaussian to the population variance of the truncated Gaussian [68].

References

[1] H. Alt, "Computational discrete mathematics, advanced lectures," Springer, 2001, Lecture Notes in Computer Science.

[2] A.A. Amini, S. Tehrani, and T.E. Weymouth, "Using dynamic programming for minimizing the energy of active contours in the presence of hard constraints," in *Proceedings for the 2nd International conference on Computer Vision*, 1988, pp. 95–99.

[3] A.A. Amini, T.E. Weymouth, and R.C. Jain, "Using dynamic programming for solving variational problems in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 9, pp. 855–867, September 1990.

[4] M.A. Anastasio, M.A. Kupinski, and R.M. Nishikawa, "Optimization and FROC analysis of rule-based detection schemes using a multiobjective approach," *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, pp. 1089–1093, December 1998.

[5] B.A. Artwick, *Microcomputer displays, graphics, and animation*, Prentice-Hall Inc., Englewoods Cliffs, New Jersey, 1st edition, 1985, pp. 25-28.

[6] S.M. Astley, C.R.M Boggis, K. Walker, S. Wallace, S. Tomkinson, V. Hillier, and J. Morris, "An evaluation of a commercial prompting system in a busy screening centre," in *Proceedings for the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 471–475.

[7] S.M. Astley and P. Miller, "Classification of breast tissue by texture analysis," *Image and Vision Computing*, vol. 10, no. 5, pp. 277–282, June 1992.

[8] S.M. Astley, T.C. Mistry, C.R.M Boggis, and V.F. Hillier, "Should we use humans or a machine to pre-screen mammograms?," in *Proceedings for the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 476–480.

[9] K. Atkinson, *Elementary numerical analysis*, John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1985, pp. 121-129.

[10] J.A. Baker, P.J. Kornguth, J.Y. Lo, M.E. Williford, and C.E. Floyd, "Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon," *Radiology*, vol. 196, no. 3, pp. 817–822, September 1995.

[11] G.T. Barnes, "Radiographic mottle: A comprehensive theory," *Medical Physics*, vol. 9, no. 5, pp. 656–667, 1982.

[12] H.H. Barrett and W. Swindell, "Theory of Random Processes", in *Radiological Imaging*, Academic Press, New York, 1st edition, 1981, pp. 62-116.

[13] J.E. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society: Series B*, vol. 48, no. 3, pp. 259–302, 1986.

[14] U. Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, and K. Doi, "Automated segmentation of digitized mammograms," *Academic Radiology*, vol. 2, no. 1, pp. 1–9, January 1995.

[15] R.E. Bird, T.W. Wallace, and B.C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology*, vol. 184, pp. 613–617, 1992.

[16] P.J. Bones, K.J. McLoughlin, and H.M. McLoughlin, "Data clustering for discrimination of microcalcifications in mammograms," in *Proceedings of the Image & Vision Computing New Zealand conference (IVCNZ)*, D.N. Kenwright, Ed., Auckland, New Zealand, November 2002, pp. 181–184.

[17] N.F. Boyd, L. Martin, G. Lockwood, C. Greenberg, M.J. Yaffe, and D. Tritchler, "Diet and breast cancer," *Nutrition*, vol. 14, pp. 722–724, 1998.

[18] R. Briant, P. Bethwaite, M. Elwood, B. Marshall, T. Green, A. Harris, P. Chapman, T. Doyle, C. Lewis, J. Sceats, and J. Simpson, "Population-based breast cancer screening: policy advice for a New Zealand screening programme," *The New Zealand Medical Journal*, vol. 111, no. 1064, pp. 138–142, April 1998.

[19] L.J.W. Burhenne, S.A. Wood, C.J. D'Orsi, S.A. Feig, D.B. Kopans, K.F. O'Shaughnessy, E.A. Sickles, L. Tabar, C.J. Vyborny, and R.A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology*, vol. 215, no. 2, pp. 554–562, May 2000.

[20] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, November 1986.

[21] D. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Medical Physics*, vol. 16, no. 4, pp. 561–568, July 1989.

[22] D. Chakraborty, "Statistical power in observer-performance studies: Comparison of the receiver operating characteristic and free-response method in tasks involving localization," *Academic Radiology*, vol. 9, pp. 147–156, 2002.

[23] D.P. Chakraborty, "The FROC, AFROC and DROC variants of the ROC analysis," in *Handbook of Medical Imaging: Physics and Psychophysics*, M. Sonka and J. M. Fitzpatrick, Eds., Chapter 16. The International Society for Optical Engineering, 2000.

[24] H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. Macmahon, "Improvement in radiologist's detection of clustered microcalcifications on mammograms," *Investigative Radiology*, vol. 25, no. 10, pp. 1102–1110, October 1990.

[25] H.P. Chan, K. Doi, S. Galhotra, C.J. Vyborny, H. MacMahon, and P.M. Jokich, "Image feature analysis and computer-aided diagnosis in digital radiography: Automated detection of microcalcifications in mammography," *Medical Physics*, vol. 14, pp. 538–548, 1987.

[26] H.P. Chan, K. Doi, C.J. Vyborny, K.L. Lam, and R.A. Schmidt, "Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study," *Investigative Radiology*, vol. 23, pp. 664–671, 1988.

[27] H.P. Chan, B. Sahiner, K.L. Lam, N. Petrick, M.A. Helvie, M.M. Goodsitt, and D.D. Alder, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Medical Physics*, vol. 25, no. 10, pp. 2007–2019, October 1998.

[28] H.P. Chan, B. Sahiner, N. Petrick, K.L. Lam, and M.A. Helvie, "Effects of pixel size on classification of microcalcifications on digitized mammograms," in *"Proceedings for SPIE Medical Imaging*, San Diego, California, February 1996, vol. 2710, pp. 30–41.

[29] H.P. Chan, C.J. Vyborny, H. MacMahon, C.E. Metz, K. Doi, and E.A. Sickles, "Digital mammography: ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications," *Investigative Radiology*, vol. 22, pp. 581–589, 1987.

[30] B. Cox, *BreastScreen Aotearoa: December 1998 to December 2000. BreastScreen Aotearoa Independent Monitoring Group Report to the Ministry of Health*, Department of Preventative and Social Medicine, Hugh Adam Cancer Epidemiology Unit, University of Otago, 33rd edition, 2002.

[31] T.S. Curry, J.E. Dowdey, and R.C. Murry, *Christensen's Physics of Diagnostic Radiology*, Lea and Febiger, Southwestern Medical Center at Dallas and Parkland Memorial Hospital, 4th edition, 1990.

[32] A.C. Davison and D.V. Hinkley, *Bootstrap methods and their application*, Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, United Kingdom, 1st edition, 1997.

[33] C. de Boor, "Piecewise Cubic interpolation", in *A practical guide to splines*, Springer-Verlag, New York, 2nd edition, 2001, pp. 39-50.

[34] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured image using Gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 39–55, January 1987.

[35] R.C. Dubes and A.K. Jain, "Random field models in image analysis," vol. 15, no. 2, pp. 131–164, 1989.

[36] R.C. Dubes, A.K. Jain, S.G. Nadabar, and C.C. Chen, "MRF model-based algorithms for image segmentation," in *Proceedings for the 10th International Conference on Pattern Recognition*, Atlantic City, USA, June 1990, pp. 808–814.

[37] R.O. Duda and P.E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Graphics and Image Processing*, vol. 15, no. 1, pp. 11–15, January 1972.

[38] B. Efron, *The jackknife, and bootstrap and other resampling plans*, J.W. Arrowsmith Ltd, Winterstoke Road, Bristol, England, 1st edition, 1982.

[39] B.S. Everitt, *Cluster Analysis*, John Wiley and Sons, New York, 3rd edition, 1993.

[40] R. Fahrig, J.A. Rowlands, and M.J. Yaffe, "X-ray imaging with amorphous selenium: Optimal spectra for digital mammography," *Medical Physics*, vol. 23, pp. 557–567, January 1996.

[41] E.C. Fear, P.M. Meaney, and M.A. Stuchly, "Microwaves for breast cancer detection?," *IEEE Potentials*, vol. 22, no. 1, pp. 12–18, February 2003.

[42] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, and F.R. Frere, "Segmentation of mammograms: Identification of the skin-air boundary, pectoral muscle, and fibroglandular tissue," in *Proceedings for the 5th International Workshop on Digital Mammography*, M.J. Yaffe, Ed., Toronto, Canada, June 2000, pp. 573–579.

[43] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Optical Society of America*, vol. 4, no. 12, pp. 2379–2394, December 1987.

[44] A. Fischmann, K. Siegmann, A. Wersebe, N. Fersis, C.D. Claussen, and M. Mller-Schimpfle, "Intraindividual comparison of full field digital mammography and conventional mammography in a breast cancer early detection program," in *Proceedings*

for the 6th International Workshop on Digital Mammography, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 481–485.

[45] J. Fraser, *Cancer, New registrations and deaths 1998*, Ministry of Health, Ministry of Health, PO Box 5013, Wellington, New Zealand, 2002.

[46] W.T. Freeman and E.H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, September 1991.

[47] T.W. Freer and M.J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology*, vol. 220, no. 3, pp. 781–786, August 2001.

[48] M.A. Gavrielides, J.Y. Lo, and C.E. Floyd, "Parameter optimization of a computer-aided diagnosis scheme for the segmentation of microcalcification clusters in mammograms," *Medical Physics*, vol. 29, no. 4, pp. 475–483, April 2002.

[49] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, November 1984.

[50] M. L. Giger, N. Karssemeijer, and S. G. Armato, "Guest editorial computer-aided diagnosis in medical imaging," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1205–1208, December 2001.

[51] M. L. Giger, F.F Yin, K. Doi, C.J. Vyborny, and R.A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral subtraction technique," *Medical Physics*, vol. 21, pp. 445–452, 1994.

[52] M. L. Giger, F.F Yin, K. Doi, Y. Wu, C.J. Vyborny, R.A. Schmidt, and Z. Huo, "Computerized detection and characterization of mass lesions in digital mammography," *Systems, Man and Cybernetics*, vol. 2, pp. 1370–1372, October 1992.

[53] M.L. Giger, K. Ohara, and K. Doi, "Effect of quantization on digitized noise and detection of low-contrast objects," in *Proceedings for the SPIE Medicine PACS conference*, Newport Beach, California, February 1986, vol. 626, pp. 214–224.

[54] R.C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley Publishing, Massachusetts, 2nd edition, 1987.

[55] S.J. Graham, M.J. Bronskill, J.W. Byng, M.J. Yaffe, and N.F. Boyd, "Quantitative correlation of breast tissue parameters using magnetic resonance and x-ray mammography," *British Journal of Cancer*, vol. 73, pp. 162–168, 1996.

[56] R. Haehnel and C. Kleitz, *Mammography: Exercises in Radiological Diagnosis*, Springer-Verlag, 1988.

[57] R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision*, Addison Wesley Publishing Company, Massachusetts, 1st edition, 1992, pp. 28-48.

[58] M. Heath, K.W. Bowyer, and D. Kopans, "Current status of the digital database for screening mammography," in *Proceedings for the 4th International Workshop on Digital Mammography*, N. Karssemeijer, Ed., Nijmegen, The Netherlands, June 1998, pp. 457–460.

[59] P. Heid, B. Sradour, C. Pagliuchi, N. Motte, S. Kurkdjian, and L. Piana, "Digital mammography vs. screen-film mammography: a phantom study," in *Proceedings for the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 20–26.

[60] P. Heinlein, J. Drexl, and W. Schneider, "Integrated wavelets for enhancement of microcalcifications in digital mammography," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 402–413, March 2003.

[61] R. Highnam and M. Brady, *Mammographic Image Analysis*, Kluwer Academic Publishers, PO Box 17, 3300 AA Dordrecht, The Netherlands, 1st edition, 1999.

[62] L.E. Hughes, R.E. Mansel, and D.J.T. Webster, *Benign disorders and diseases of the breast: Concepts and clinical management*, W.B. Saunders Harcourt Publishers Limited, 1st edition, 1989.

[63] L.E. Hughes, R.E. Mansel, and D.J.T. Webster, *Benign disorders and diseases of the breast: Concepts and clinical management*, W.B. Saunders Harcourt Publishers Limited, 2nd edition, 2000.

[64] Z. Huo, M.L. Giger, C.J. Vyborny, U. Bick, P. Lu, D.E. Wolverton, and R.A. Schmidt, "Analysis of spiculation in the computerized classification of mammographic masses," *Medical Physics*, vol. 22, no. 10, pp. 1569–1579, October 1995.

[65] A.K. Jain, *Fundamentals of digital image processing*, Prentice Hall, Englewoods Cliffs, New Jersey, 1st edition, 1989.

[66] A.K. Jain, M.N. Murty, and P.J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, September 1999.

[67] Y. Jiang, R.M Nishikawa, D.E. Wolverton, C.E. Metz, M.L Giger, R.A. Schmidt, C.J. Vyborny, and K. Doi, "Malignant and benign clustered microcalcifications: automated feature analysis and classification," *Radiology*, vol. 198, no. 3, pp. 671–678, March 1996.

[68] N.L. Johnson, S. Kotz, and N. Balkrishnan, *Continuous univariate distributions*, John Wiley and Sons Inc, New York, 2nd edition, 1994, pp. 156-162.

[69] T.D. Jones and P. Plassmann, "An active contour model for measuring the area of leg ulcers," *IEEE Transactions on Medical Imaging*, vol. 17, no. 12, pp. 1202–1210, December 2000.

[70] N. Karssemeijer, "Images were provided by courtesy of the National Expert and Training Centre for Breast Cancer Screening and the Department of Radiology at the University of Nijmegen, The Netherlands," .

[71] N. Karssemeijer, "A stochastic model for automated detection of calcifications in digital mammograms," *Image and Vision Computing*, vol. 10, no. 6, pp. 369–375, July 1992.

[72] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clustesr in mammograms," vol. 7, no. 6, pp. 1357–1375, 1993.

[73] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," *Physics of Medicine and Biology*, vol. 43, pp. 365–378, 1998.

[74] N. Karssemeijer, "Personal correspondence," 2003.

[75] N. Karssemeijer, J.T.M Frieling, and J.H. Hendricks, "Spatial resolution in digital mammography," *Investigative Radiology*, vol. 28, no. 5, pp. 413–419, May 1993.

[76] N. Karssemeijer and J.H. Hendricks, "Computer assisted reading of mammograms," *European Radiology*, vol. 7, no. 5, pp. 743–748, 1997.

[77] N. Karssemeijer and G.M. te Brake, "Detection of stellate distortions in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 611–619, October 1996.

[78] N. Karssemeijer and L. van Erning, "Iso-precision scaling of digitized mammograms to facilitate image analysis," *SPIE Image Processing*, vol. 1445, pp. 166–177, 1991.

[79] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, pp. 321–331, 1988.

[80] W.P. Kegelmeyer, "Computer detection of stellate lesions in mammograms," in *Proceedings of the 1992 SPIE Conference on Biomedical Image Processing and 3-D Microscopy*, San Jose, California, February 1992, vol. 1660, pp. 446–454.

[81] W.P. Kegelmeyer, J.M. Pruneda, P.D. Bourland, A. Hillis, M.W. Riggs, and M.L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology*, vol. 191, no. 2, pp. 331–337, May 1994.

[82] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

[83] H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 235–245, June 1996.

[84] J.J. Koenderink and A.J van Doorn, "Generic neighborhood operators," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 597–605, June 1992.

[85] D.B. Kopans, *Breast Imaging*, Lippincott-Raven, Philadelphia, 2nd edition, 1998.

[86] P. Kovesi, "Image features from phase congruency," *Videre: A Journal of Computer Vision Research. MIT Press*, vol. 1, no. 3, pp. 1–26, 1999.

[87] M.A. Kupinski and M.L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 510–517, August 1998.

[88] S.M. Kwok, R. Chandrasekhar, and Y. Attikiouzel, "Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection," in *Proceedings for the 7th Australia and New Zealand Intelligent information systems conference*, 2001, pp. 67–72.

[89] A. Laine, J. Fan, and W. Yang, "Wavelets for contrast enhancement of digital mammography," *IEEE Engineering in Medicine and Biology*, pp. 536–550, September 1995.

[90] L.M. Lam and H. Yan, "Locating head boundary by snakes," in *Proceedings for the International Symposium on Speech, Image Processing and Neural Networks*, Hong Kong, April 1994, pp. 17–20.

[91] F. Lee, L.C. Yee, and S. Bressan, "Registration of maps using deformable contours in global atlas," in *Proceedings for the 12th International Workshop on Database and Expert Systems Applications*, Munich, Germany, September 2001, pp. 875–879.

[92] G. Lemaur, K. Drouiche, and J. DeConinck, "Highly regular wavelets for the detection of clustered microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 393–401, March 2003.

[93] H.D. Li, M. Kallergi, L.P. Clarke, V.K. Jain, and R.A. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 565–576, September 1995.

[94] S.P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, March 1982.

[95] C.S. Lo, P.C Chung, S.K. Lee, C.I. Chang, T. Lee, G.C. Hsu, and C.W. Yang, "Off-line mammography screening system embedded with hierarchically-coarse-to-fine techniques for the detection and segmentation of clustered microcalcifications," *IEEE Transactions on Information and Systems*, vol. E83-D, no. 12, pp. 2161–2173, December 2000.

[96] A.D.A. Maidment, R. Fahrig, and M.J. Yaffe, "Dynamic range requirements of x-ray detectors for digital mammography," *Medical Physics*, vol. 20, pp. 1621–1633, 1993.

[97] A.D.A. Maidment and M.J. Yaffe, "Analysis of signal propagation in optically-coupled detectors for digital mammography: II. Lens and fibre optics," *Physics of Medicine and Biology*, vol. 41, pp. 475–493, 1996.

[98] A. Maitournam, C. Graffigne, and A. Stauss, "Modeling of digital mammograms using bicubic spline function and additive noise," in *Proceedings for the 4th International Workshop on Digital Mammography*, N. Karssemeijer, M. A. O. Thijssen, J. H. Hendriks, and L. J. van Erning, Eds., Nijmegen, The Netherlands, June 1998, pp. 177–184.

[99] R.E. Mansel, Ed., *Recent developments in the study of benign breast diseases*, Parthenon Publishing Group, 4th edition, 1990.

[100] M. Masek, R. Chandrasekhar, C.J.S. deSilva, and Y Attikiouzel, "Spatially based application of the minimum cross-entropy thresholding algorithm to segment the pectoral muscle in mammograms," in *Proceedings for the 7th Australia and New Zealand Intelligent information systems conference*, 2001, pp. 101–106.

[101] MathWorks, "Worldwide Headquarters, The MathWorks, Inc. 3 Apple Hill Drive, Natick, MA 01760-2098, USA," .

[102] K.J. McLoughlin, P.J. Bones, and P.D. Kovesi, "Detection of microcalcifications in digital mammograms," in *Proceedings of the Image & Vision Computing conference (IVCNZ)*, K. Novins and B. McCane, Eds., Dunedin, New Zealand, November 2001, pp. 259–264.

[103] K.J. McLoughlin, P.J. Bones, and P.D. Kovesi, "Connective tissue representation for detection of microcalcifications in digital mammograms," in *Proceedings of the 2002 SPIE Conference on Medical Imaging*, M. Sonka and J.M. Fitzpatrick, Eds., San Diego, California, February 2002, vol. 4684, pp. 1246–1256.

[104] K.J. McLoughlin, P.J. Bones, and P.D. Kovesi, "Detection of microcalcification clusters in digital mammogram images," in *Proceedings of the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 353–357.

[105] K. McPherson, C.M. Steel, and J.M. Dixon, "ABC of breast cancer: Epidemiology, risk factors, and genetics," *British Medical Journal*, vol. 321, pp. 624–628, September 2000.

[106] A. Mendez, P. Tahoces, M. Lado, M. Souto, J. Correa, and J. Vidal, "Automatic detection of breast border and nipple in digital mammograms," *Computer Methods and Programs in Biomedicine*, vol. 49, pp. 253–262, 1996.

[107] A.J. Mendez, P.G. Tahoces, M.J. Lado, M. Souto, and J.J. Vidal, "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms," *Medical Physics*, vol. 25, no. 6, pp. 957–964, June 1998.

[108] C.E. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology*, vol. 21, no. 9, pp. 720–733, September 1986.

[109] C.E. Metz, "Evaluation of digital mammography by ROC analysis," in *Proceedings of the 3rd international workshop on digital mammography*, K. Doi, M.L. Giger, R.M. Nishikawa, and R.A. Schmidt, Eds., Chicago, June 1996, pp. 61–68.

[110] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory - part II: Sampling theory and complex waves.," *Geophysics*, vol. 47, no. 2, pp. 222–236, February 1982.

[111] M.C. Morrone and R.A. Owens, "Feature detection from local energy," *Pattern Recognition Letters*, vol. 6, pp. 303–313, 1987.

[112] M.C. Morrone, J.R. Ross, D.C. Burr, and R.A. Owens, "Mach bands are phase dependent," *Nutrition*, vol. 324, no. 6094, pp. 250–253, November 1986.

[113] R.H. Nagel, R.M. Nishikawa, J. Papaioannou, and K. Doi, "Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms," *Medical Physics*, vol. 25, no. 8, pp. 1502–1506, August 1998.

[114] H. Neiber, T. Muller, and R. Stotzka, "Local contrast enhancement for the detection of microcalcifications," in *Proceedings for the 5th International Workshop on Digital Mammography*, M.J. Yaffe, Ed., Toronto, Canada, June 2000, pp. 598–604.

[115] T. Netsch and H.O. Peitgen, "Scale-space signatures for the detection of clustered microcalcifications in digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 18, no. 9, pp. 774–786, September 1999.

[116] R.M. Nishikawa, "Detection of microcalcifications," in *Image processing techniques for tumor detection*, R.N. Strickland, Ed., Chapter 6. Marcel Dekker Inc., 2002.

[117] R.M. Nishikawa, M.L. Giger, K. Doi, C.J. Vyborny, and R.A. Schmidt, "Computer-aided detection of clustered microcalcifications: An improved method for grouping detected signals," *Medical Physics*, vol. 20, no. 6, pp. 1661–1666, November 1993.

[118] R.M. Nishikawa, Y. Jiang, M.L. Giger, K. Doi, C.J. Vyborny, and R.A. Schmidt, "Computer-aided detection of clustered microcalcifications," *Systems, Man and Cybernetics*, vol. 2, pp. 1375–1378, October 1992.

[119] R.M. Nishikawa and M.J. Yaffe, "Signal-to-noise properties of mammographics film-screen systems," *Medical Physics*, vol. 12, no. 1, pp. 32–39, February 1985.

[120] N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, January 1979.

[121] S.C. Pei and J.H. Horng, "Determination of circular arc length and midpoint by Hough transform," in *Proceedings for the 3rd International symposium on circuits and systems*, June 1994, pp. 1–4.

[122] R. Pfisterer and F. Aghdasi, "Comparison of texture based algorithms for the detection of masses in digitised mammograms," *IEEE AFRICON*, vol. 1, pp. 383–388, October 1999.

[123] S. Pohlman, K.A. Powell, N.A. Obuchowski, W.A. Chilcote, and S. Grundfest-Broniatowski, "Quantitative classification of breast tumors in digitized mammograms," *Medical Physics*, vol. 23, no. 8, pp. 1337–1345, August 1996.

[124] M. Poissonnier and M. Brady, "Noise equalization, film-screen artifacts, and density representation," in *Proceedings for the 5th International Workshop on Digital Mammography*, M.J. Yaffe, Ed., Toronto, Canada, June 2000, pp. 605–611.

[125] W. Qian, M. Kallergi, L. P. Clarke, and R. A. Clark, "Tree-structured nonlinear filters in digital mammography," *IEEE Transactions on Medical Imaging*, vol. 13, no. 1, pp. 25–36, March 1994.

[126] W. Qian, M. Kallergi, L. P. Clarke, H. D. Li, P. Venugopal, D. Song, and R. A. Clark, "Tree structured wavelet transform segmentation of microcalcifications in digital mammograms," *Medical Physics*, vol. 22, pp. 1247–1254, August 1995.

[127] F. Richard and L. Cohen, "Non-rigid mammogram registration with free boundary constraints," in *Proceedings for the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 213–217.

[128] A. Richardson, B. Cox, and P. Graham, "Should there be a breast cancer risk chart for New Zealand women?," *The New Zealand Medical Journal*, vol. 112, no. 1086, pp. 129–130, April 1999.

[129] H. Roehrig, E. Krupinski, and T. Yu, "Physical and psychophysical evaluation of digital systems for mammography," in *SPIE Proceedings on Medical Imaging: Image Perception*, H.L. Kundel, Ed., San Jose, February 1995, vol. 2436, pp. 124–134.

[130] B. Sahiner, H.P. Chan, N. Petrick, L.M. Hadjiiski, M.A. Helvie, and S. Paquerault, "Active contour models for segmentation of characterization of mammographic masses," in *Proceedings for the 5th International Workshop on Digital Mammography*, M.J. Yaffe, Ed., Toronto, Canada, June 2000, pp. 357–362.

[131] P.K. Sahoo, S. Soltani, and A.K.C. Wong, "A survey of thresholding techniques," *Computer Vision, Graphics and Image Processing*, vol. 41, pp. 233–260, 1988.

[132] P. Sajda, C. Spence, and J. Peason, "Learning contextual relationships in mammograms using a hierarchical pyramid neural network," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 239–250, March 2002.

[133] V.U.B. Schenk and M. Brady, "Finding CLS using multiresoltuion oriented local energy feature detection," in *Proceedings for the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 64–68.

[134] F. Schmidt, E. Sorantin, C. Szepesvari, E. Graif, M. Becker, H. Mayer, and K. Hartwagner, "An automated method for the identification and interpretation of clustered microcalcifications in mammograms," *Physics of Medicine and Biology*, vol. 44, pp. 1231–1243, 1999.

[135] G.F. Schwartz and D. Marchant, Eds., *Breast Disease: Diagnosis and Treatment*, Symposia Specialists Inc, Philadelphia, 1981.

[136] M. I. Sezan, "A peak detection algorithm and its application to histogram-based image data reduction," *Computer Vision, Graphics and Image Processing*, vol. 49, pp. 36–51, 1990.

[137] E.P. Simoncelli and H. Farid, "Steerable wedge filters for local orientation analysis," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1377–1382, September 1996.

[138] D. Skegg, C. Paul, D. Benson-Cooper, J. Chetwynd, A. Clarke, N. Fitzgerald, I. St George, and A. Simpson, "Mammographic screening for breast cancer: prospects for New Zealand," *The New Zealand Medical Journal*, vol. 101, no. 852, pp. 531–533, August 1988.

[139] J. Sklansky, "On the Hough technique for curve detection," *IEEE Transactions on Computers*, vol. 27, no. 10, pp. 923–926, October 1978.

[140] J.A. Smallwood and I. Taylor, Eds., *Benign breast disease*, Edward Arnold, Division of Hodder and Stoughton, University Surgical Unit, SouthHampton, 1990.

[141] R.N. Strickland and H.I. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, no. 2, pp. 218–229, April 1996.

[142] R.N. Strickland and H.I. Hahn, "Wavelet transform method for object detection and recovery," *IEEE Transactions on Image Processing*, vol. 6, no. 5, pp. 724–, May 1997.

[143] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "The mammographic image analysis society digital mammogram database (MIAS)," *Exerpta Medica. International Congress Series 1069*, vol. 1069, pp. 375–378, 1994.

[144] G.M. te Brake and N. Karssemeijer, "Automated detection of breast carcinomas not detected in a screening program," *Radiology*, vol. 207, pp. 465–471, 1998.

[145] G.M. te Brake, M.J. Stoutjesdijk, and N. Karssemeijer, "A discrete dynamic contour model for mass segmentation in digital mammograms," in *Proceedings of SPIE Medical Imaging: Image Processing*, K. M. Hanson, Ed., San Diego, California, February 1999, vol. 3661, pp. 911–919.

[146] S. Timp, N. Karssemeijer, and J.H Hendriks, "Comparison of three methods for segmentation of masses in mammograms," in *Proceedings for the 6th International Workshop on Digital Mammography*, H.O. Peitgen, Ed., Bremen, Germany, June 2002, pp. 218–222.

[147] G.T. Toussaint, "The relative neighbourhood graph of a finite planar set," *Pattern Recognition*, vol. 12, pp. 261–268, 1980.

[148] W.H. Tsai, "Moment-preserving thresholding: A new approach," *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 377–393, 1985.

[149] M. Unser, "Splines, a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 99, pp. 22–38, November 1999.

[150] W. Veldkamp and N. Karssemeijer, "Normalization of local contrast in mammograms," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 731–738, July 2000.

[151] C.J. Vyborny, "Can computers help radiologists read mammograms?," *Radiology*, vol. 191, no. 2, pp. 315–317, 1994.

[152] T.C. Wang and N.B. Karayiannis, "Detection of microcalcifications in digital mammograms using wavelets," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 498–509, August 1998.

[153] D.J. Williams and M. Shah, "A fast algorithm for active contours," in *Proceedings for the 3rd International conference on computer vision*, 1990, pp. 592–595.

[154] K. Woods, M.Y. Sallam, and K.W. Bowyer, "Evaluating detection algorithms," in *Image processing techniques for tumor detection*, R.N. Strickland, Ed., Chapter 2. Marcel Dekker Inc., 2002.

[155] C.H. Wu, Y.C. Chen, and K.S. Hseih, "Texture features for classification of ultrasonic liver images," *IEEE Transactions on Medical Imaging*, vol. 11, no. 2, pp. 141–152, June 1992.

[156] Y. Wu, M.L. Giger, K. Doi, C.J. Vyborny, R.A. Schmidt, and C.E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer," *Radiology*, vol. 187, no. 1, pp. 81–87, April 1993.

[157] M.J. Yaffe, "Digital mammography," in *Handbook of Medical Imaging: Physics and Psychophysics*, M. Sonka and J. M. Fitzpatrick, Eds., Chapter 5. The International Society for Optical Engineering, 2000.

[158] S. Yu and L. Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 115–126, February 2000.

[159] R. Zwiggelaar, T.C. Parr, J.E. Schumm, I.W. Hutt, C.J. Taylor, S.M Astley, and C.R. Boggis, "Model based detection of spiculated lesions in mammograms," *Medical Image Analysis*, vol. 3, no. 1, pp. 39–62, March 1999.